The Validity And Reliability Of Value-Added And Target-Setting Procedures With Special Reference To Key Stage 3

Thesis

IAN ROBIN MOODY

P0126273

THE VALIDITY AND RELIABILITY OF VALUE-ADDED AND TARGET-SETTING PROCEDURES WITH SPECIAL REFERENCE TO KEY STAGE 3

DOCTOR OF EDUCATION (EdD)

2003

For Palmerston,

my consolation,

in dark moments

ACKNOWLEDGEMENTS

In any large scale project there are likely to be times of triumph and times of despair. In a project lasting three years, motivation tends to be very high at the outset but tends to be increasingly difficult to sustain as the years pass by. In order to see such a project through to conclusion, therefore, the support, interest, enthusiasm, concern and empathy of others are often of inestimable importance in retrospect. Furthermore, whilst new technology provides us with opportunities for efficiency and presentation undreamt of twenty years ago, it also provides us with the potential for inducing high levels of anxiety when it fails, or even just threatens to fail. My greatest debt in all these respects is to Jenifer, Jean and Terry, whose motivation and interest, at times, seemed to far outstrip my own. Without their support, completion of this project would have been a very lonely journey. I would like to thank Anne, David, Duncan, Helena, Jane, Mercedes, Rosemary and Zona who gave freely of their time amidst hectic professional lives, to take part in the staff interviews. Their honesty, candour, and highly perceptive insights into the issues surrounding value-added and target-setting were immensely valuable. Professor David Scott, who supervised this project was inspirational. His detailed critiques of various drafts, his scholarship, the breadth of his ideas and insights, and above all his understanding of the very human difficulties which surround such a project were of the very highest order. He was always available and willing to talk ideas through, often at very short notice. Whatever merit others may deem this work to have, arises substantially, from his outstanding attributes, both as a teacher and as a scholar.

CONTENTS

LIST OF TABLES

LIST OF APPENDICES

So little trouble do men take in the search after truth;

So readily do they accept whatever comes first to hand.

Thucydides: *The Peloponnesian War*

ABSTRACT

The validity of value-added systems of measurement is crucially dependent upon there being a demonstrably unambiguous relationship between the so-called baseline, or intake measures, and any subsequent measure of performance at a later stage. The reliability of such procedures is dependent on the relationships between these two measures being relatively stable over time. A number of questions arise with regard to both the validity and reliability of value-added procedures at any level in education, but this appears to be particularly problematic at Key Stage 3. Target-setting procedures usually employ value-added data as the basis for predicting future performance, and the validity and reliability of target-setting procedures is therefore also problematic. In a five-year longitudinal case-study of one secondary comprehensive school, the validity of both of these procedures was investigated using quantitative and qualitative methods. The validity and reliability of Key Stage 2 data, and data from the NFER Cognitive Abilities Test (CAT), as baseline data for value-added procedures, was investigated using regression analysis. Wider perceptions of value-added procedures were opened up using detailed interviews with a representative cross-section of teachers. Target-setting procedures were investigated by analysing numerical performance data in curriculum areas, by the use of questionnaires to students, staff and parents, and by detailed interviews with a representative sample of students. The results of the investigation suggest that whilst value-added procedures have a high level of validity and reliability for Mathematics at Key Stage 3, this is less true for English and Science. For non-core subjects there appears to be no rational basis upon which value-added procedures can be used. Staff attitudes towards both value-added and target-setting were often quite ambivalent; those of students and parents were generally much more positive. A number of concerns emerged with regard to the ways in which such procedures appear to be distorting teaching and learning, producing a culture of performativity where learning is increasingly seen as a means to

an end, rather than a worthwhile end in itself. These findings raise some fundamental questions about our philosophy of education, about accountability, and about the knowledge claims which are made as a result of such procedures.

CHAPTER 1

CONTEXT OF THE STUDY

NATIONAL CONTEXT

Anxiety about educational standards, like anxiety about the moral standards of the young, is both perennial and deep-rooted in modern society. Sociologists describe periods of intense concern about moral standards as times of "moral panic" and a similar phenomenon is evident with regard to concerns about educational standards. In Britain, this kind of deep-rooted anxiety about educational standards became apparent during the 1970s (Chitty, 1994) This anxiety was given a very clear focus in a speech by James Callaghan, the then Labour Prime Minister, at Ruskin College, Oxford, in 1976. Callaghan had asked the DES to provide him with information about various aspects of primary and secondary education, particularly about standards in reading, writing and arithmetic. The so-called "Yellow Book", which the DES produced, motivated Callaghan to call for a national debate about educational standards and the purposes of education. As in the American context, there were deeper fears and anxieties driving these concerns, not least the rapid rise in unemployment, rapidly deteriorating industrial relations, and an increasing sense that the UK was declining as an industrial power in the modern world.

Callaghan (1976) identified "complaints from industry" (p.10) about the standards of new recruits from schools, a lack of "desire to join industry" (p.11) by graduates, "unease felt by parents and others about the new informal methods of teaching" (p.12) and the need for "a basic curriculum with universal standards" (p.13). Callaghan went on to raise questions about "the proper way of monitoring the use of resources in order to maintain a proper national standard of performance" (p.19), about the examination system and staying on rates after the age of sixteen, and about the governance of schools. Somewhat remarkably, having raised such

fundamental questions, Callaghan claimed that he did not want to "join those who paint a lurid picture of educational decline because I do not believe it is generally true" (p.17). Whether he believed this or not, it triggered an intense sense of national anxiety about the state of education, the effects of which are still very much with us twenty-six years later, as we will see. With hindsight, what is clear, is that this was a significant turning point in our education system, not least because it was evident from proposals in the "Yellow Book" that moves were afoot to give both the DES and HMI much greater powers over a wide range of educational matters. Phillips (2001), for instance, has documented how, by 1997, the education system was "unrecognisable from the one that existed in 1976" (p.19) and that by 1997 "all aspects of education were closely controlled by government" (p.19). The two proposed developments which we can now see as fundamental to these changes were the issue of a core curriculum for schools, and the issue of how schools could be made more accountable for their standards of performance. The response of the *Times Educational Supplement* (TES 1976) was that "the immediate effect was one of anticlimax" (p.1). Yet, almost paradoxically, it went on to describe Callaghan as using "weasel words to exploit popular prejudices" (p.1) by trotting out "cliches from the CBI about the shortcomings of young workers" (p.1). The paper was firmly of the opinion that since the economy had been in decline for fifteen years, the real problem was the need to "change working conditions and human relations at the factory level" (p.2). The fundamental changes needed in society, it opined, could not be achieved by "tinkering with educational influences" (p.2) because this was "like inviting water to flow up hill" (p.2).

Sir Alec Clegg (1976), writing in the same edition of the TES, pointed out that an earlier Chief Inspector of schools at the beginning of the twentieth century, Edmond Holmes, had decided to institute a core curriculum so that "children could be compared with children, teachers with teachers, and schools with schools" (p.2) and that upon retirement the same Inspector

wrote a book "in which he condemned lock, stock and barrel, all that he had been doing for 30 years (p.2). Drawing on Ruskin, Clegg expressed concerns about the idea of educating students for industry, asking "are we to teach solely what the building up of the gross national product demands and devil take the hindmost?" (p.2).

One outcome of this deep-rooted anxiety over educational standards had been the *Black Paper* which launched repeated attacks, during the 1970s, on what it perceived to be the failings of the education system. In two editorials in the *Black Paper 1977*, the leaders of progressive education were described as combining "good intentions with ignorance and vain hope" (p.12) and the editors concluded that "all the accumulating evidence indicates that formal structured traditional teaching methods are superior to informal non-structured education" (p.19). The editors attacked policies at all levels, including LEAs, government, teacher training institutions and schools themselves. They saw Callaghan as having "attempted to steal our clothes, which have always been freely available" (p.5). He was seen as repeating "our assertions that money is being wasted, standards are too low, and children are not being given the basic tools of literacy and numeracy" (p.6). In 1969, when the *Black Paper* was first launched, according to the editors, the paper was "treated with abuse and contempt" (p.6) but, as they saw it, "the evidence that standards were declining" had become "overwhelming" (p.6).

Another outcome of this anxiety over educational standards, both here, and in the USA, was the "school effectiveness" movement. The landmark study which is generally regarded as establishing the movement in the UK was *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children* (Rutter et al 1979). This study aimed to identify the characteristics of effective schools with a view to other, less effective schools, learning from them. It set the agenda for the kinds of questions which school effectiveness research would be asking for the next twenty years. The central concern in this type of research was to establish how much of a difference (if any) a

school makes to the individual who passes through it. This "intake-process-outcome" model was seen as representing a fundamentally new way of comparing the effectiveness of individual institutions. By taking account of intake ability, and comparing it with outputs, it was assumed that it was possible to compare schools on a fairer basis, and identify what it was that apparently made some schools much more effective than others. The companion study to *Fifteen Thousand Hours* was *School Matters* (Mortimore et al 1988) which examined the effectiveness of primary schools. Taken together, the two studies claimed to be able to identify a number of characteristics of effective schools and as Mortimore (1998) noted subsequently, it was hoped that this research would inform debate and that the dissemination of such findings would lead to the development of more effective schools. Critics have asked whether it is possible, in practice, to transplant characteristics of effective performance from one institution to another but Mortimore (1998) remains convinced that "all that is needed is for policy makers to listen to the whole message and to have the courage to heed it" (p.345).

As the 1980s progressed it was clear that a much more radical philosophy was emerging with regard to raising educational standards, and national anxiety was intensified by constant attacks on the teaching profession by central government. The new philosophy was founded on a number of assumptions, which to some extent drew on earlier ideas, but it took these ideas a great deal further than had ever been envisaged before. One assumption was that government needed to establish a standardised curriculum, and standardised forms of national assessment and testing (to enable comparisons to be made between schools). Another was that the public should be given as much information as possible about the performance of individual schools. A third, potent idea, arising from the idea of publishing performance data about schools, was to give parents complete freedom about where they wanted their children to be educated. In a nutshell, the idea was that information about schools, together with the

effects of parents being able to choose which school their child attended, would subject schools to the kinds of market forces which business and industry has to respond to, and that this would, over time, drive up standards (Phillips 2001).

The 1988 Education Reform Act embodied all of these ideas. Drawing on the recommendations of TGAT (1988), a common, ten-subject curriculum would be compulsory for all students aged five to fourteen (with some choice available for the fourteen to sixteen age group). Each subject would have ten levels of attainment, enabling progress over time to be compared. Students would be tested, nationally, at the ages of seven, eleven and fourteen, and the results of these tests, together with the results of the new GCSE examination, would be published in the form of school league tables. The concept of league-tables, clearly drawn from the world of professional football, again embodied this idea that publicly available information, showing where any school was in the hierarchy of success or failure, would drive up standards. The 1988 Act fundamentally changed the balance of power in education and, as Coulby (1989) has noted it "shifted huge areas of control over educational institutions to the Secretary of State at the DES" (p.3).

Whilst the new national tests at seven, eleven and fourteen provided a solid basis upon which national comparisons of schools could be made, a new set of problems quickly emerged. One was the bureaucracy that this kind of accountability gives rise to, threatening to undermine the very improvements it is designed to bring about (Broadfoot 2001). Secondly, the publication of league tables rapidly revealed what many critics had been forecasting: that position in the national league tables would largely reflect the kind of area in which a school was situated. Schools in areas of disadvantage, not surprisingly, tended to find themselves much lower in the league tables. The analogy with football league tables is instructive in this respect. Just as poor football clubs do not have the money to build better facilities, pay the kind of wages which enable them to attract more able

players, and therefore attract bigger crowds, so schools in poor areas found themselves in a competitive struggle in which there often appeared to be little that they could do to improve their lot. A third problem was that the original idea that all ten subjects should be tested at seven, eleven and fourteen quickly turned out to be unworkable, and national testing never developed beyond the three core subjects of English, Mathematics and Science. Progress in the other seven National Curriculum subjects is based entirely on teacher assessments. Given that these assessments are not nationally moderated, the consistency of these judgements is somewhat open to question (QCA 1998) and this creates significant problems in any attempt to measure progress in these subjects.

The fundamental problems posed by the idea of league tables were addressed by two new methods of comparing educational standards nationally in the mid-1990s. The first of these was "benchmarking". Each school, from late 1997 onwards, was provided with an annual report, the Autumn Package, and the Performance and Assessment (PANDA) Report by OFSTED, which attempted to compare schools on the basis of their *intake*, rather than on the basis of their *raw examination results*, as league tables had done. Schools were put into various benchmark groups depending on what proportion of their students were eligible for free school meals. The DFES takes this to be a reasonable indicator of differing levels of social deprivation, though it is not without its problems as OFSTED was to find, and as Goldstein (2001) has shown (see Chapter 3). Within each benchmark group, schools were then placed in one of four quartiles within that group, depending on their results. Whilst this new method of making comparisons appeared to be fairer, it soon became apparent that schools from the same benchmark groups were often achieving widely different kinds of standards in national tests. The problem is, of course, that defining schools in terms of their social intake (measured in terms of the percentage of students entitled to free school meals) does not necessarily tell you very much about the prior attainment, or *ability*, of those students. Whilst PANDA reports did use

prior attainment to measure progress at Key Stage 4 (using Key Stage 3 average test scores), it failed to use this procedure at Key Stage 3. As the school effectiveness movement had suggested, intake ability, and prior attainment, are highly significant if one is attempting to measure how well a school is performing. The new method which emerged, in an attempt to address this problem, was therefore "value-added".

Value-added is based on the idea that you can measure prior attainment, or ability, when a student enters a school, measure attainment again when they leave the school, and then, by comparing it with the performance of students of similar intake ability, or attainment in other schools, or with historical data of the same kind, make some estimate of how much "value" the school has added. This procedure was recommended by TGAT (1988) as the basis of the original National Curriculum assessments. Despite recognising that such a system of educational measurement was extremely problematic, SCAA (1997a), in *The Value Added National Project Final Report* recommended that a national system of value-added should be established. At Key Stage 3, it recommended that value-added should be calculated from 1998 using Key Stage 2 test levels to predict outcomes at Key Stage 3. As will become evident in Chapter 3, this approach is fraught with problems, not the least of which is that, despite the original intention of the 1988 Act that the ten levels of attainment across each subject would enable the system to identify how much progress any student had made, in practice serious questions began to emerge about the lack of equivalence between levels at Key Stage 2 and Key Stage 3. Put simply, if Level 4 at Key Stage 2 is not the same as Level 4 at Key Stage 3, it is difficult, if not impossible, to calculate progress in any meaningful way, let alone compute how much value has been added. Furthermore, since value-added depends on making predictions about outcomes, based on historical data, it is necessary for the data to remain stable and consistent over time. As Moody (2001) has shown, in a case study of the first two intake years upon which this study is based, this assumption cannot be taken for granted. This raises quite fundamental

questions about the very foundations upon which the concept of value-added is built. Furthermore, as we will see in Chapter 3, the answers you get in value-added calculations depend on the *methods* you use, and *what you put into the model*. As Goldstein (2001) has shown, the amount of value any given school can be "shown" to have added can vary enormously, depending on the assumptions you make about what data should, or should not, be included.

Given the inadequacies of each of these new developments in trying to measure national and local standards in education, and given also the immense bureaucracy that this has spawned, both inside and outside schools, it might have seemed sensible to put a stop to such developments. However, rather than learning from these problems, government seemed determined to intensify the use of such procedures. The next idea was to give schools performance targets. The DFEE (1996) argued that schools needed to be made part of the process of achieving the national targets for education and training, which had already been established following pressure from the CBI in 1991. The basic philosophy of this report was that targets help schools to raise their standards by clarifying goals and by placing a "greater onus on pupils to do better" (p.6). SCAA (1997b) recommended that schools should be required to set performance targets from September 1998, adding that "many of the most effective schools succeed in raising pupils' attainment, sometimes quite dramatically, by setting targets for improvement based on how their pupils are currently performing, and taking specific action to meet these targets" (p.1). Schools were to be required to publish their targets for performance in public examinations at ages sixteen and eighteen and, subsequently, to publish information about the extent to which they had achieved these targets.

Setting targets, of course, involves being able to make some kind of prediction about where a school might expect to be, two or three years into the future, a notoriously difficult exercise even given valid and reliable data upon which to make such predictions. However, evidence was already in the

public domain which suggested that much of the data that was available to make such predictions was neither valid nor reliable. Wiliam (1993), in a study of the validity, dependability and reliability of National Curriculum assessment concluded that "government claims that National Curriculum assessments will be both reliable and valid cannot be taken seriously" (p.348) and that the government was "establishing reliability and validity by fiat" (p.348). Yet SCAA (1997b) felt able to advise the government that "National Curriculum assessment now provides consistent and reliable measures of pupils' attainments at ages 7, 11 and 14 and will therefore form the basis for the targets schools will set" (p.4). It could be argued, of course, that four years had elapsed between Wiliam's conclusions and the recommendations of SCAA, but as this study reveals, there are still serious questions to be answered about the validity and reliability of national test results at Key Stages 2 and 3, even in 2001. ATL (2000), commenting on the proposals for target setting at Key Stage 3, concluded from their studies of the Key Stage 3 tests that they were "not yet of a suitably high quality to be used for target setting purposes" (p.3) and that the urgency of the government in rushing into this development needed to be "moderated by a sober analysis of the appropriateness and the quality of the instruments being used for measuring attainment" (p.3). Nevertheless the DFEE (2001) subsequently required schools to set targets at Key Stage 3 for national test results in the summer of 2002. This proposal arose as a result of what became known as the "Key Stage 3 dip". National data tended to suggest that whereas pupils were beginning to make significant progress over their years at primary school, between the ages of eleven and fourteen they often made little or no progress (DFEE 2000). In the white paper *Schools Achieving Success* (DFES 2001), the government claimed that "the greatest immediate challenge is to transform standards in the first years of secondary school" (p.17) because "there has been growing concern about the slow progress that children make between the ages of 11 and 14" (p.17).

An even more controversial element in this process was to be imposed on the profession in 2001, namely an element of performance-related pay, dependent on meeting specified targets. This had been exactly what the profession had feared the government was moving towards. Having established a pay "threshold" which teachers had to apply to cross, in order to access a higher pay scale, teachers could then only progress further up this higher pay scale if they met agreed targets annually. One of these targets each year was to be related to pupil progress. Many members of the profession found this move extremely disturbing because, like all the other developments outlined so far, it is bedevilled by problems which are very difficult, if not impossible, to resolve. Given the problems we have already encountered with regard to using assessment data to raise standards, the idea of using such data for accountability purposes, and for making decisions about the pay of individual teachers, raises some fundamental questions about fairness and equity (Goldstein 2001).

So, twenty six years after Callaghan's speech where are we? In a speech celebrating the twentieth anniversary of the Callaghan Speech, Blair (1996), then leader of the opposition, reported Callaghan as having observed that "our problems today are more urgent than they were then" (p.3) and supported this with a series of statistics about the education system which suggested that he was correct in this assertion. In 2001, four years after the Labour party came to power, the white paper cited in the previous paragraph (DFES 2001) inevitably tried to put an optimistic gloss on the state of the education system, but some of the government's own statements in this paper suggest that there were still substantial problems in terms of educational standards. The paper suggested that "as a country we are still wasting an enormous amount of talent by denying some of our children the quality of education that would make a real difference to their lives" (p.4), that "in 1997 nearly half of 11 year olds were below standard in basic literacy" (p.5) and that even in 2000, slightly less than 50% of students nationally were achieving five or more A*-C grades at GCSE. At Key

Stage 3 "for too many pupils, the first year or two of secondary school can be a time of falling motivation and rising disaffection" (p.13), and that as a result the performance of fourteen year olds between 1998 and 2000 has shown "relatively little improvement" (p.13). In post-compulsory education "we are well down the OECD international league tables for pupils staying on in education beyond the age of 16" (p.14). In fairness, there are statistics which show marked improvements in standards, such as the rise between 1996 and 2000 in the number of students achieving Level 4 in English at Key Stage 2, from 57% to 75%, and the 5% rise in the same period in the percentage of students achieving five or more GCSE's at grades A*-C. Tymms and FitzGibbon (2001) have raised serious questions, however, about the validity of these supposed improvements, especially in Key Stage 2 English (discussed at in more detail in Chapter 6), The white paper also reported a significant drop in the number of primary school lessons rated during OFSTED inspections as unsatisfactory between 1995 and 2000, from 17% to 4%. In secondary schools, for the same period, the figures were 16% and 6%. Yet across the system as a whole, 138 schools were still in special measures in 2000-2001 as a result of poor performance. The government has therefore set ambitious targets to be achieved at the end of Key Stage 3 by 2007, with 85% of students being expected to achieve Level 5 or above in English, Mathematics and ICT, and 80% to have achieved Level 5 or above in Science. In order to drive up standards further, value-added procedures are to be used in published data (starting with a pilot in 2001), and for the first time national test results for fourteen year olds will be published too.

LOCAL CONTEXT

This study was undertaken in a secondary comprehensive school for girls aged 11-19 (with a mixed sixth form). In the early 1980s this had been a very successful school with high levels of academic achievement, but by the early 1990s the school was in serious decline in terms of academic

achievements, behaviour, the physical state of the buildings and resources generally. Recruitment, in the free market created by the Conservative governments of the 1980s, was falling rapidly, and by 1993 the school was facing a situation where the potential intake in Year 7 could have been as low as eighty-five students. Repeated year on year, this would have meant that the school would no longer have been viable within five years and would have closed. In a city where there had been a large number of surplus places for several years, there was a suspicion within the school that the LEA would not attempt to intervene if it appeared to be heading for closure, because it would save the authority from having to take a difficult decision which it had been trying to avoid for years.

However, under new management from 1993, the school began to turn itself around and by 2001 was in the position where it was so popular that it was over-subscribed and was having to turn students away in Year 7. The statistics shown below (Table 1.1 – 1.4) demonstrate how dramatic this change was, particularly between 1997 and 2001. The school adopted a whole series of initiatives to raise standards, drawing on school effectiveness research, on the use of value-added procedures, and more recently on individual target-setting procedures for students. Staff who worked in the school showed very high levels of commitment and determination, and the effects of this can be seen in the statistics below, particularly in the benchmark comparisons. Of particular interest is performance at Key Stage 3. Whereas, as we have noted, there has been increasing national concern about the lack of progress at Key Stage 3, in this school there was a rise in achievement at Key Stage 3.

The school has a genuinely "comprehensive" intake, which draws students from local authority housing estates, from middle-income owner occupied areas, and from families with parents in high-status professional jobs. The city has a large number of private schools which achieve very high academic standards. It is an interesting commentary on the intake of the school that many of the parents of students who attend the school in which

this study took place, openly admit that whilst they *could* have afforded to send their daughters to a private school, they *chose* not to. The school population therefore represents a very interesting cross section of society.

Staffing within the school has remained relatively stable over time in terms of staff turnover, though throughout the mid-1990s overall staffing levels were dropping due to budget cuts. However, in the summer of 2001 some 25% of the staff decided to leave. Whilst some of these were taking retirement, or going to other schools, a number of others had decided to leave teaching, either for other jobs, or to be unemployed. As interview evidence gathered for this study reveals (see Chapter 5), to some extent this reflected the enormous pressures which many staff felt themselves to be under in the school, but a number had also become increasingly disillusioned with national policies on education and no longer felt that they believed in what they were doing. The bureaucracy associated with value-added and target-setting, and the associated distorting effects on the curriculum, and teacher-pupil relationships, was seen by many staff who left the school as having reached a point beyond which they were not prepared to go. The introduction of performance management, and performance-related pay, had only served to deepen the sense of disillusion for many of these staff.

Statistical data about the school where this study was conducted provide a useful overview of the context for the study, particularly the comparisons with national averages. The data was extracted from the PANDA reports for the school for the years 1997-2001. The cohorts of students who were the focus of this study were the intake years 1994-1998. These students sat their Key Stage 3 tests between 1997 and 2001, the years to which these data apply (see Tables 1.1 – 1.6).

TABLE 1.1

BASIC CHARACTERISTICS OF THE SCHOOL IN WHICH THE STUDY WAS UNDERTAKEN

	1997	1998	1999	2000	2001
Number on roll					
School	914	946	957	1006	1099
National Average	907	915	934	955	983
Percentage of pupils eligible for free school meals					
School	N/A	N/A	13.2	13.3	11.5
National Average	N/A	N/A	18.1	17.8	15.3
Percentage of pupils speaking English as a second language					
School	2.8	1.4	1.9	1.8	0.8
Percentage of pupils with special educational needs (including statements)					
School	21.1	16.5	17.2	16.4	9.2
National average	18.9	17.9	18.7	19.3	19.5
Percentage of students with statements of special educational needs					
School	1.5	1.5	1.3	1.4	1.2
National average	2.3	2.4	2.5	2.5	2.5

Table 1.1 shows how the numbers on roll had increased by some 20% over the period, reflecting the success in recruitment mentioned earlier. The numbers of students speaking English as a second language has dropped over the period, as has the number of students with special educational needs. The number of students eligible for free school meals has consistently been below the national average, but has been high enough to affect the benchmark group into which the school is placed by the DFEE.

TABLE 1.2

AVERAGE POINTS SCORE AT KEY STAGE 3

	1997	1998	1999	2000	2001
English					
School	32.6	34.5	34.3	34.8	35.6
Girls nationally	33.4	34.7	33.9	34.0	34.4
School difference (girls)	-0.8	-0.2	0.4	0.8	1.2
All pupils nationally	32.0	33.2	32.5	32.6	33.0
School difference (all pupils)	0.6	1.3	1.8	2.2	2.6
Mathematics					
School	35.1	34.8	34.8	34.9	35.3
Girls nationally	33.0	32.8	33.4	34.2	34.5
School difference	2.1	2.0	1.4	0.7	0.8
All pupils nationally	33.2	33.1	33.5	34.3	34.4
School difference (all pupils)	1.9	1.7	1.3	0.6	0.9
Science					
School	34.1	33.3	32.8	32.6	33.5
Girls nationally	32.4	31.5	31.5	31.9	33.1
School difference (girls)	1.7	1.8	1.3	0.7	0.4
All pupils nationally	32.6	31.8	31.6	32.2	33.1
School difference (pupils)	1.5	1.5	1.2	0.4	0.4

Table 1.2 shows that the school has been achieving above the national average at Key Stage 3 for all three subjects. However, whereas in English the trend has been to perform at a level increasingly higher than the national average, in Mathematics and Science, the gap between the school's performance, and national performance is closing. The reasons for the rising levels of performance in English, compared with national performance is not entirely clear, though English staff often indicated that rising levels of literacy at Key Stage 2 enabled them to progress students much more quickly at Key Stage 3.

TABLE 1.3

COMPARISON OF ATTAINMENT WITH ALL SCHOOLS NATIONALLY

	1997	1998	1999	2000	2001
Key Stage 3 National Curriculum tests (average points)					
English	B	B	A	A	A
Mathematics	A	A	B	B	B
Science	B	B	B	C	C
All subjects	B	B	B	B	B
GCSE/GNVQ					
5 or more grades A*-C	C	C	B	B	B
5 or more grades A*-G	D	C	D	C	B
1 or more grades A*-G	D	C	E	B	D
Average total GCSE points score per pupil	C	C	B	B	B
GCE A/AS level					
Average points score for pupils entered for 2 or more A levels or AS equivalent	C	D	C	C	B

Note 1

Grades are awarded in PANDA reports on the following scale in relation to the national average:

A* = very high A = well above B = above

C = broadly in line D = below E = well below

E* = very low

Table 1.3 clearly illustrates the extent to which the school is achieving above, or well above, the national average at Key Stage 3 (with a slight decline in Science in 2000 and 2001). At GCSE the picture is generally one of improvement over the period, whilst at A/AS level, the school has held its own against the national average.

TABLE 1.4

COMPARISON OF ATTAINMENT WITH SCHOOLS IN THE SAME BENCHMARK GROUP

	1997	1998	1999	2000	2001
Key Stage 3 National Curriculum tests (average points)					
English	N/A	N/A	A	A	A
Mathematics	N/A	N/A	A	A	B
Science	N/A	N/A	A	A	C
All subjects	N/A	N/A	A	A	B
GCSE/GNVQ					
5 or more grades A*-C	N/A	N/A	A	A	B
5 or more grades A*-G	N/A	N/A	D	B	B
1 or more grades A*-G	N/A	N/A	E	A	D
Average total GCSE points score per pupil	N/A	N/A	A	A	B

Note 1

The grades are based on the same scale as for Table 1.3 except that the comparison is with the *average for similar schools*. (The benchmark group for the school in 1999 and 2000 was non-selective maintained schools with more than 13%, and up to 21%, free school meals. For 2001 the benchmark group was non-selective maintained schools with more than 9% and up to 13% free school meals).

Table 1.4 shows just how well the school has performed in terms of benchmark comparisons with similar schools, both at Key Stage 3 and at GCSE/GNVQ. The school was moved into a more demanding benchmark group in 2001, as a result of falling numbers of students who were eligible for free school meals. This demonstrates some of the potential problems schools face with regard to evaluation of their performance, as their population changes over time, particularly where a change to another benchmark group is marginal.

TABLE 1.5

ATTENDANCE RATE (%)

	1996/7	1997/8	1998/9	1999/0	2000/1
School	93.4	93.5	92.3	91.9	91.8
National average (secondary schools)	90.9	91.0	91.0	93.7	90.9
School difference	2.5	2.5	1.3	-1.8	0.9

The school has made considerable efforts with regard to maintaining attendance levels and Table 1.5 shows that this has generally been very successful, maintaining attendance above the national average.

TABLE 1.6

PRIOR ATTAINMENT OF STUDENTS ENTERING YEAR 7 1994-1998 AS MEASURED BY NFER COGNITIVE ABILITIES TEST

	Mean Standard Age Score	Standard Deviation
1994 intake		
Verbal	100	15.0 (not significant)
Quantitative	96	13.9 (not significant)
Non-verbal	103	13.1 (significant)
1995 intake		
Verbal	100	13.5 (significant)
Quantitative	96	12.2 (significant)
Non-verbal	103	11.9 (significant)
1996 intake		
Verbal	100	14.4 (not significant)
Quantitative	97	12.9 (significant)
Non-verbal	102	11.7 (significant)
1997 intake		
Verbal	98	14.6 (not significant)
Quantitative	95	12.6 (significant)
Non-verbal	99	12.9 (significant)
1998 intake		
Verbal	99	14.6 (not significant)
Quantitative	98	13.1 (significant)
Non-verbal	100	13.0 (significant)

Note 1

The mean standard age score nationally for each sub-test is 100.

Note 2

The standard deviation nationally for each sub-test is 15.0.

Note 3

The 1994 intake group sat test Level F at the beginning of Year 9. The 1995 intake sat test Level E at the beginning of Year 8.The intakes for 1996-1998 sat test Level D at the beginning of Year 7. Since scores on the tests are age-standardised by the NFER, theoretically the scores should be comparable.

Table 1.6 illustrates how the intake for each year has achieved a mean standard age score close to the national average. The spread of these results however, as shown by the standard deviation, is often significantly different from the national average, suggesting that student ability is not normally distributed. However, scores on the Verbal test show no significant difference, except for the 1995 intake, suggesting a normal ability distribution on this element of the tests.

The school started using the A-Level Information System (ALIS) in the mid 1990s to estimate value-added from GCSE to A level. In 1998 the author was appointed as the Assessment Manager for the school in which this study was undertaken, with a brief to extend the use of value-added procedures throughout the school, and to establish a target-setting and tracking procedure for individual students at Key Stages 3 and 4, and at A/AS level. After extensive research into target-setting procedures, in schools which were already pioneering this approach to raising standards, the school decided to evaluate this evidence and make a decision about which procedure to employ. There were a number of alternatives in use in other schools. Some schools used baseline data to predict *likely* outcomes at Key Stage 3 and Key Stage 4 and then set what is known as a *target minimum grade* for each student. Tracking procedures within these schools were designed to ensure that no student fell below this level in any subject. Some

schools set a target for *each* subject, whilst others set *one target* for all subjects across the board. Some schools were even setting a number of sub-targets within the overall subject target.

Experience gained from contact with schools using these methods of setting targets suggested that having a large number of targets rapidly gave rise to an extremely bureaucratic process, especially if the school wished to track the performance of individual students. It was also clear that students were often unable to hold all of these targets in their heads anyway, largely defeating the purpose of having targets in the first place. Schools which had gone a step further and set up sub-targets within each subject were clearly finding themselves overwhelmed by the bureaucracy involved. For schools employing the target-minimum-grade approach there was evidence that once students achieved this minimum target they tended to lose their motivation. The school therefore decided to set up what came to be known as high challenge targets. The idea was to use the baseline data to predict what each student might reasonably be expected to achieve at the end of a Key Stage, and then set an average target across all subjects, one level, or one grade, higher than this. The calculated target was reported to the student, to staff, and to the parents, as a *High Challenge Target Average*, and the initial estimation was that about a third of students in a given subject area would achieve this target.

The plan was that a brief interim report would be issued in the autumn term of each year, which would report progress towards this target for each subject. This was to represent a prediction about what each student would be likely to achieve at the *end of the Key Stage*, on the basis of current performance. In order to make this process more meaningful for students, teachers and parents, the aim was to move away from the traditional, verbally-based forms of reporting, such as "satisfactory" or "good", and establish a clear and more definitive judgement about progress. Each student would therefore receive a report which would identify whether they were predicted to achieve the target at the end of the Key Stage, be slightly

below, or above the target, or be one, two, or three grades above or below the target. This form of reporting progress was also used as part of a second, longer report, later in the year. Parents were given information about how these procedures had been established, and the concept of progress towards targets became part of the process of consulting with parents when they visited the school.

At Key Stage 3 a number of models were tested using baseline data from the NFER (1986) Cognitive Abilities Test (CAT), in order to establish suitable targets for students. Experience suggested that Key Stage 2 test data were very unreliable, but in any case it was often not arriving in the school until the spring term, and the data for quite a large number of students never arrived at all. After testing a range of models the school decided to use the average CAT score, and set the target on the basis of the scores shown in Table 1.7. (A more detailed discussion of the CAT, and its use, can be found in Chapter 3).

TABLE 1.7

PROCEDURE FOR SETTING KEY STAGE 3 TARGETS

Average CAT score	Target level at Key Stage 3
70-85	4
86-99	5
100-110	6
111-129	7
130+	8

These targets came into operation for all students at Key Stage 3 in the academic year 1999-2000 and for all subjects except Art, Music and Physical Education (PE). For the latter three subjects, targets were introduced in the academic year 2000-2001 when the revised National Curriculum had established clearer levels of attainment in these subjects than had existed previously.

At Key Stage 4 the decision was taken to use the average test level at Key Stage 3 in English, Mathematics and Science. The LEA provided all schools with so-called "chances tables", based on the average Key Stage 3 test level. These types of data give percentage chances for any student at a particular level at Key Stage 3, of achieving any particular grade at Key Stage 4.

Again, students were given a High Challenge Target Average for all subjects. Table 1.8 shows how this worked out in practice.

TABLE 1.8

PROCEDURE FOR SETTING KEY STAGE 4 TARGETS

Average Key Stage 3 test score	**High Challenge Target Average Grade for all GCSE subjects**	**Percentage of students likely to achieve this grade or higher in any GCSE subject**
Up to 3.99	E	37.6%
4.00 to 4.59	D	40.1%
4.60 to 5.29	C	35.5%
5.30 to 5.99	B	24.9%
6.00 and above	A	28.2%

At A level (and from 2000 AS level) the target was calculated by taking the ALIS predicted grade and setting the target one grade higher. Students were given individual targets for each subject.

AUTOBIOGRAPHICAL CONTEXT

I entered the teaching profession in the year of Callaghan's speech, 1976, after education at a selective grammar school (entry to which was dependent on passing the 11+ examination), university, and post-graduate secondary teacher training. I have twenty-six years experience in four secondary schools and one further education college, and ten years experience as a part-time lecturer in higher education.

Having attended a school with extremely high academic standards, I was shocked, upon entering the profession, to discover how poor standards often were in state comprehensive schools. Whilst many teachers were clearly doing a very good job, and many students were achieving high standards, it was also clear that there was a great deal in the state education system that needed attention. I therefore identified with some of the concerns raised in Callaghan's speech, particularly with regards to standards in reading, writing and arithmetic. In these early years in the profession it appeared that many schools in the maintained sector essentially had it within their power to be achieving much higher standards. On the other hand, the funding of schools did not seem to be very good, and the general state of books and equipment often left a lot to be desired. Support from LEAs too, often appeared to be somewhat perfunctory, whilst in terms of building maintenance, support not only seemed to be very poor, but often seemed to betray a history of long-term neglect.

The attacks on the profession, originating from government during the early 1980s, became increasingly critical, but despite my own view that schools could be doing a lot more to raise standards, I increasingly found myself losing sympathy with the attitudes emanating from government, not least because, as noted above, I certainly did not take the view that the system as a whole was failing. Whilst there was clearly room for improvement, and in some notorious cases, room for radical improvement, it was certainly not the case that the entire profession was ineffective, although this was often what the public might have supposed, listening to various Secretaries of State for Education during the 1980s. It seemed to me that what was needed, fundamentally, was to recognise where individual schools and teachers were doing well, to celebrate this, and to win the support of those within the profession who were dedicated to raising standards. Instead, what successive governments seemed to be doing was to denigrate the work of the profession as a whole, and create high levels of public anxiety about the work of all schools, and all teachers. By 1987, as Coulby (1989) notes,

"teachers were defeated, discredited and demoralised" (p.10) and realised that "their power and discretion were aspects of the system that future legislation would endeavour to curb" (p.10). Blair (1996), commenting on these years of Conservative government, suggested that "one of the government's biggest mistakes has been to take the excesses of a few teachers as an excuse to pillory the whole profession" (p.6). The removal of pay negotiating rights in the mid 1980s and the subsequent passing of the 1988 Act were, therefore, significant turning points for the profession. Whilst I could see some merits in a common curriculum, and national testing which would enable comparisons to be made year on year, it seemed unlikely that government centralising so much control of education would be a healthy development in the long run. The problems associated with the publication of league tables only served to highlight these concerns, and the subsequent developments in the 1990s in value-added, benchmarking, target-setting, and performance management, with all their shortcomings have, in my view, done a great deal of damage to education.

The issue for me, therefore, is not *whether* we should be collecting data in this way, and making comparisons in order to try to raise standards. The issue is the *way in which such data is being used*, and its effects on schools, teachers, students, parents, and the perceptions of the general public about what is really going on in education. The problems associated with assessment are legion. Add to this the potential for complex statistical data to be manipulated and presented in a variety of ways which are, at the very least, potentially misleading, and the ingredients are in place for considerable conflict and misunderstanding, both inside and outside the profession. Such conflict and misunderstanding, too often, directs the energies of schools and teachers away from the very business of raising standards which these procedures are supposed to be designed to bring about.

Value-added is a problematic concept generally, but it faces a number of additional difficulties if it is applied at Key Stage 3. Two issues, in

particular, arise. One is whether Key Stage 2 data is sufficiently valid and reliable to be used as baseline data upon which to make value-added calculations at Key Stage 3, and if not, whether data such as the NFER (1986) Cognitive Abilities Test provides a more valid and reliable baseline. The other issue is whether the Key Stage 3 test results, and teacher assessments, are reliable enough to make the kind of year-on-year comparisons which are necessary in value-added procedures.

Since target-setting is founded on the kinds of data analysis which value-added procedures give rise to, the issue of how valid and reliable target-setting procedures are, is worthy of further investigation, not least because targets have very obvious consequential effects on teachers and students. As we will see in Chapter 3 with the work of Messick (1993), the *consequences* of any form of educational measurement are a crucial element in the validity of any measurement. It is important, therefore, that we investigate, and try to understand better, what effects such targets are having on the way the education system is working at Key Stage 3.

This study therefore set out to investigate three, broad questions:

1 What insights can be gained with regard to the validity of value-added and target-setting procedures at Key Stage 3?

2 What insights can be gained with regard to the reliability of value-added and target-setting procedures at Key Stage 3?

3 What are the philosophical and practical implications of these insights?

SUMMARY

Assessment has become one of the most contested areas in education, with regard to both the procedures used to gather such data, and the kinds of knowledge claims which can be sustained about such data. Since value-added and target-setting procedures involve making comparisons and

predictions, these concepts are clearly even more problematic. This study set out to investigate these issues in terms of existing practice in one local school (described earlier). Investigation of such procedures necessitates being very clear about the types of research methodology which are potentially available, and about the potential validity of any knowledge claims which may be made as a result of employing any of these methods. It is these issues which we will examine in some depth, therefore, in Chapter 2.

CHAPTER 2

METHODOLOGICAL ISSUES AND RESEARCH RATIONALE

PREAMBLE

Saint John, in his Biblical account of the trial of Christ, reports how Pontius Pilate, confronted with Christ's claim to truth, asks "what is truth?" (John 18:38). The question, and Pilate's dilemma, ring out across the millenia. Confronted with a conflict between the evidence of our own perceptions and cognitions, and those of others, especially if they are in a majority, how do we choose? Do we accept that the majority always has a stronger claim to truth than a minority, or even an individual? Can a single individual be right, and an entire organisation or society be wrong? Is there any such thing as objective, knowable truth, or is truth whatever we perceive it to be? If there is some kind of objective, knowable truth, how do we know where the boundaries of this truth actually lie? Are these boundaries ever clear-cut? Whilst Pilate's question appears to represent a reaction to Christ's spiritual claims for himself, and appears to be largely rhetorical or speculative, it is clear from the succeeding account that Pilate's question also represented a much more practical, down to earth, immediate problem: faced with a situation where we *have to do something*, where we *have to make a decision one way or the other*, where a decision is *quite inescapable*, on what basis do we decide? Are there any certainties upon which we can base our decisions, or are they ultimately always driven by a choice between a stark series of potentially unreliable and subjective alternatives such as pragmatism, expediency, self-interest, ideology, a belief in some kind of revealed truth, or even just blind fatalism?

WHAT CAN BE KNOWN? EVIDENCE FROM PHILOSOPHY

The history of epistemology, across many different cultures, and more than two millenia, reveals how philosophers have struggled to achieve some kind of certainty about what can, and what cannot be known. The ideas which have been put forward have increasingly left us with more questions than answers. Plato (1993), for instance, nearly two and half thousand years ago, characterised the experience of most individuals as being like prisoners in a cave, chained facing a wall, where all they can ever see is the shadows of a reality outside the cave entrance behind them, and about which they can never have any direct knowledge. All they can do is to generate a set of *beliefs*. True knowledge of the reality outside the cave is possible for a few who can rise out of the cave, the so-called philosopher-kings, but as Cooper (1999) has pointed out, it remains a matter of debate as to whether such philosophers can return to the cave and offer any insights to the prisoners which would make any kind of intelligible sense to them, or indeed whether the prisoners themselves would see such knowledge as having any relevance to their existence or experience. Plato's view suggests that only a small elite can ever know reality, and understand, and explain, what is going on in human experience, whilst the experience of the majority is to grope around in a twilight of speculation and supposition. The idea that truth can only be known by the few obviously appears to us, today, to be fundamentally problematic, not least because of the implications with regard to concepts such as power-relations, privilege, and control in a democratic society.

Aristotle's view (1999) of what can be known arises from his conception of *episteme*, or what we might now call scientific knowledge or understanding. Aristotle's work established the epistemological tradition known as foundationalism where knowledge is deducible from a series of basic principles or premisses. However, the issue of how we come to know these premisses is unclear because Aristotle does not believe that these principles are demonstrable. Whilst he accepts that what we could now call perception

and cognition play a part in how we come to know such premisses, ultimately we only know them as a result of intuition (Cooper 1999). This suggests that we have some kind of innate mechanism for discerning truth from falsity. A number of questions arise from this assumption. Do all individuals have this mechanism? Is the mechanism equally reliable in everybody? How do such intuitions arise? Are they some kind of revealed truth? Are the intuitions of all individuals equally valid? If they are, how do we resolve differences in individual intuitions?

The tradition of scepticism, established by Pyrrho and his followers in classical Greece and discussed by Sextus Empiricus (1999) is, perhaps, the epistemological tradition which has had the most lasting influence on our thinking right down to the present. Scepticism essentially refuses to accept that any one set of beliefs about reality is preferable to its rivals. This comes about as a result of a suspension *of* judgement, and a refusal to *make* judgements. In terms of trying to cope with the ever present conflicts that we face in our experience with regard to what can be known with any certainty, scepticism could appear to have considerable appeal, at least in individual, philosophical terms. In our modern, knowledge-rich society, for instance, where we are bombarded with conflicting assertions, often claiming to be empirically-based, it is clearly very tempting to suspend judgement. That way we avoid having to make decisions which may well change the course of our actions. In the world of *practical* realities, of course, as Pilate found, suspension of judgement is rarely possible for very long, because the pressure of events around us has an unfortunate tendency to place us in a position where we not only *have* to make judgements between competing sets of claims but, even more problematically, we then have to *act* on our conclusions, and accept the consequences which inevitably follow.

If we accept that we are indeed driven by the pressure of events, ultimately, to make judgements, the question is *how* we make those judgements, and *on*

what basis. Are they merely arbitrary? Whose opinion should we trust? Can we afford to trust one opinion more than another? Do we put more trust in those who lay claim to some sort of authority (say as a result of expert knowledge)? Do we trust those who hold appointed positions of authority in society? Do we assume that those in positions of authority, or who lay claim to expert knowledge, are always, to a greater or lesser extent, compromised by the power-structures within which they operate? Are those in positions of authority being *allowed* to tell us the truth as they see it? Are opinions supported by empirical evidence, based on widely-accepted methods of data-collection, to be trusted? We see it as the mark of an open, democratic society that conflicting claims can be made available for public debate, discussion and evaluation but this leaves us with a whole series of dilemmas which it appears to be very difficult, if not impossible, to resolve without retreating into some kind of fatalism, ideology, or belief in revealed truth.

These various traditions, evident in classical Greek philosophy, can be seen to play themselves out, in one form or another, both in the philosophy of cultures very different from our own, and also in philosophical writing within the western tradition, right down to the present day. In the period roughly contemporary with classical Greece, for instance, the sceptical tradition can be seen in the writings of Chuang Tzu (1996) in China, whilst a form of foundationalism can be seen in the Indian Nyaya Sutras (1999). In western philosophy, Descartes arrived at the conclusion that certain kinds of knowledge are rationally deducible, a position completely rejected Locke (1961), the founder of empiricism, who believed that all of our knowledge is founded in, and ultimately derived from, experience. This experience arises from our sensations of the world, and from our perception of the operation of our own mind. This suggests that the brain is somehow passively imprinted with knowledge of the world and that there might therefore be no reason why two individuals should have different kinds of knowledge about the world, which they patently do.

According to Russell (1979), for Berkeley, the material world *only* exists in our perceptions and this increasingly suggested that our knowledge of the world was entirely subjective. Kant (1991) attempted to rescue philosophy from the kind of relativism which this view of knowledge and experience was inevitably leading to, by re-establishing the idea that objective knowledge of the world is possible, through the use of our powers of cognition and reason, though whether we can ever escape subjective influences on our cognition and reason remains questionable. Husserl (1964), in his theory of phenomenology, claimed that it was possible to purge our cognitive powers of influences which distort our knowledge of the world, and thereby attain a purer apprehension of reality. By the time Husserl was writing, however, other ideas were being advanced which would ultimately lead to a complete collapse of the idea that there is any kind of objective reality which can be known. Nietzsche (1999), for instance, underlined the extent to which our conceptions of the world are imprisoned in, and limited by, the constraints of language. With the assertion of Wittgenstein (1969) that all knowledge claims are context-bound, because of their dependence on language, which is itself highly context-bound, the final death knell appeared to have been sounded for any remaining claims for an objective reality, outside of ourselves, which we can know.

WHAT CAN BE KNOWN? EVIDENCE FROM THE NATURAL SCIENCES

The emergence of what we now call the natural sciences, as part of the Enlightenment thinking of the eighteenth century, was founded on the idea that we could rationally, and objectively, investigate and measure the world around us, and as a process of coming to understand it better, begin to formulate laws about the world around us, and therefore make predictions about future outcomes. Scientific method was therefore built on the assumption that there is a reality "out there" which can be known. This so-

called "correspondence theory of truth" seemed to offer the promise to eighteenth and nineteenth century society that we could establish some kind of knowledge which was incontrovertibly "true" because it was founded in demonstrably objective, empirical methods of investigation, and rational hypothetico-deductive methods which were transparent and open to challenge. This idea came to be known as positivism, and as we will see, it had important consequences for the social sciences.

An increasingly powerful element in this scientific method was the concept of reductionism – the idea that rational forms of objective investigation are capable of identifying the basic underlying causes for the world being the way that it is. Reductionism operates on the assumption that simple cause-effect relationships underpin complex systems and by identifying these relationships, the system as a whole can be understood, outcomes predicted, and that variables can be manipulated within the system in order to produce different types of outcomes. Cohen and Stewart (1994) describe this process as equipping us with a "variety of mental funnels, with complexities at the top, deeper simplicities below" (p.180). This "structure of nested funnels provides a chain of logical explanation that leads in the reverse direction, 'upward' from simple laws to complicated features of the natural world" (p.180). However, as Cohen and Stewart point out, whilst reductionism may suggest more mathematical levels of predictability at lower levels in the system, "it gradually becomes more fuzzy as we ascend to the more complex levels" (p.180).

Reductionism was dealt a serious blow by investigations in quantum physics. Reductionist principles laid down by Newton had suggested that there are certain clear and incontrovertible laws which underlie, explain, and can predict, the realities that we see around us. What quantum physics revealed was that at the sub-atomic level the "world is governed by the accumulation of outcomes of essentially random "choices" of possibilities at the quantum level" (Gribbin 1995 p.3). The idea that the "reality" in which

we find ourselves is essentially underpinned by chaotic and unpredictable behaviours was something which Einstein was unable to accept and led to his famous assertion that "God does not play dice". Within the world of quantum physics, nothing is real unless it is observed, but unfortunately the moment we begin to observe sub-atomic particles, we change the behaviour of those particles, because if just one single photon of light enters the space which is under observation, it changes the behaviour of the other particles in that space. The dilemma, of course, is that without light we cannot see what is going on. Furthermore it is a "cardinal rule of quantum mechanics that *in principle* it is impossible to measure precisely certain pairs of properties, including position/momentum, simultaneously" (Gribbin 1995 p.120). An example of this is the so-called Heisenberg Uncertainty Principle which shows that the more accurately we know the position of an object at the sub-atomic level, the less certain we are of its momentum, and vice-versa. The inevitable conclusion therefore is that "there is no absolute truth at the quantum level" (Gribbin 1995 p.120).

As if this isn't enough of a problem for knowledge claims in the natural sciences, Cohen and Stewart (1994) have shown that whilst chaos theory reveals that "simple laws can have very complicated – indeed, unpredictable consequences" (p.2), paradoxically complexity theory suggests that "complex causes can produce simple effects" (p. 2). The problem for science is how, apparently, "great simplicities of nature can persist within a chaotic universe" (p. 2). These twin themes of chaos and complexity "lie at the frontiers of modern thinking" (p.3) because "the reductionist story is nowhere near as complete as it appears to be" (p.3). What this means is that science is "trying to come to grips with 'emergent phenomena' – collective behaviour of a system that somehow transcends its components" (p.3). What all this suggests, as we saw in quantum theory, is that we may have to accept that the natural sciences will increasingly present us not with an apparently rational and logical explanation of the world in which we find

ourselves, but with a series of paradoxes. The search for "truth" and "certainty" may be an eternal illusion.

Kuhn (1996), in his study of scientific revolutions, dealt a further blow to this idea of "truth" and "certainty" in the natural sciences with his concept of the "paradigm shift". Kuhn argued that over a period of time, the natural sciences become locked into a particular way of understanding the world and that breaking out of this "paradigm" is extremely difficult because it involves challenging fundamental assumptions. This is clearly tends to be resisted, because if our fundamental assumptions turn out to be incorrect we are faced with having to rethink our entire conception of the world around us. The idea that we may have built substantial bodies of "knowledge" on an illusion, or a misunderstanding, is obviously not something that we welcome. However Kuhn argues that within the history of science problems, contradictions and questions begin to emerge within a particular orthodoxy, which eventually reach a point where a paradigm shift is inevitable. Einstein's work on relativity theory was a classic example, and we have seen the fundamental crisis for science which was triggered by quantum research.

WHAT CAN BE KNOWN? EVIDENCE FROM OTHER DISCIPLINES

If philosophy leads us into a world of extreme relativism where nothing can ever be known with any certainty, and empirical science finds itself unable to sustain our belief in its ability to provide us with the kinds of certainties that it once did, where does this leave us? Do other disciplines suggest a way forward? The short answer often appears to be "no". Evidence from a range of other disciplines tends to lend support to the idea that our knowledge of the world is subjectively constructed, although two pieces of more recent research, as we will see, *may* provide the basis for a return to

the idea that there are at least some potentially rational certainties upon which we can build our knowledge of the world.

The dream of Psychology, at the end of the 19th century was to use objective research methods to investigate behaviour, and to establish laws of behaviour in much the same way as science had established laws of nature. Pavlov's work suggested that animals could be conditioned to behave in various pre-determined ways by altering existing stimulus-response mechanisms in the animal. Picking up on this idea, Watson (1930) claimed that if he could control the environment of a child he could take any human offspring and using conditioning techniques, turn the child into anything that he chose, doctor, artist, lawyer or whatever. Although psychology subsequently retreated some way from these more radical behaviourist claims, the work of Skinner (1938), using operant conditioning techniques was fundamentally more influential, especially in education. According to operant conditioning theory, by manipulating the environment, and the rewards and punishments which people find themselves subjected to, we can fundamentally control and influence human behaviour. Skinner claimed, as a result of his work with rats that it was possible to establish laws of behaviour associated with operant conditioning and that these "schedules of reinforcement" can be used to influence and alter behaviour. As we will see, Skinner also experimented fairly extensively with the use of "teaching machines" in a bid to improve what students can be conditioned to do in a classroom.

Koestler (1989), in a biting attack on behaviourism described it as "a ratomorphic view of man" (p.17) because it stripped human beings of any faculties beyond those of animals, on whom behaviourist theories had originally been tested. It was, in his view, an outrage on the very discipline in whose name it was originated because, by denying the very concept of *mind*, "it represented a demonstrative act of self-castration" (p.17). Research in psychology after Skinner lends a great deal of support to Koestler.

Behaviourists had taken no account of cognition, let alone the concept of the individual ego. Their deterministic assertions were inevitable enough, given the dominant positivist paradigm from which behaviourism arose, but the growth of interest in cognition from the 1960s onwards gradually undermined the kinds of knowledge and insights which behaviourism had claimed to be able to access. As we will see, however, the underlying psychology of value-added and target-setting procedures in education reveals a worrying tendency to return to the positivist assumptions of behaviourism.

Research into perception suggests that incoming information from our five senses is not *recorded* but is *interpreted*. A significant mechanism in this interpretive process is our schemas (Bartlett 1932). Schemas are mental representations of the world which we have built up from past experiences. These schemas influence what we attend to, how we perceive things that we attend to, and how we remember those perceptions. According to Bartlett we try to make sense of what is going on around us, and in seeking for some kind of meaning, we tap into our existing conceptions of how the world is, especially when we are trying to make sense of an unusual or unfamiliar situation.

Cohen (1993) suggests, therefore, that when we store information in memory, we ignore inputs from the environment which are not compatible with our current schemas, that we make inferences about what is happening based on our existing schemas, that we even distort experiences to make them fit with our schemas, and that schemas influence what we retrieve from memory. Recent experimental work by French and Richards (1993) and by Loftus (1993) tends to confirm these conclusions. Furthermore, investigations into the mechanisms underlying perception suggest that these mechanisms may, to quite a large extent, be culturally and environmentally determined, according to Segall et al (1963) and Turnbull (1961). Gestalt theories of perception suggest that the whole is always greater than the sum

of the parts – in other words that things which we observe have "emergent properties" as a result of the way that we organise and interpret the information. This suggests that our personal "knowledge" of the world around us is likely to be fairly idiosyncratic. Even more fundamental, perhaps, is the writing of Goleman (1998) on self-deception. Goleman's argument is that self-deception is necessary for our very psychological self-preservation. Self-deception enables us to keep our view of ourselves, and our view of the world, intact and we therefore have a strong tendency to ignore experience which threatens to undermine these perceptions.

Marx (1971) suggested that our consciousness of ourselves and the world around us is not something that we generate independently of our social experience, but that it is largely *determined* by it. This produces a kind of false consciousness where, instead of seeing ourselves and the world around us the way it really us, we see it in terms of various ideological influences on us from the society in which we live. Furthermore, we are largely unaware of this state of affairs.

Gombrich (1977), in his work on the perception of the visual arts, suggests that there is no such thing as an "innocent eye". As viewers of art we always bring a host of experiences and perceptions to bear on a painting or a sculpture, and this influences the way that we react to the piece, the way that we interpret the piece, and the value that we place on it. In effect, works of art, like the texts of Barthes (1974), take on a life of their own once they are finished. It is likely that there are also strong elements of our thinking about our social identity at work here too. What we choose to pay attention to, the intellectual position that we choose to take up, how we want to be seen by others, the individuals or groups that we want to identify with, or stand in opposition to, and perhaps even the desire just to be different, all play their part in our reactions to the world around us at various times.

One attempt which was made to try to establish a rational basis upon which we might found our epistemological assumptions was by Popper (1972). Popper suggests that what we have to do is to treat all knowledge as provisional. Knowledge progresses by a series of conjectures and refutations. Criticisms of conjectures highlight errors, mistaken assumptions, and so on, and this enables us to understand the problem better. In this way our knowledge and understanding grows and develops, but we can never reach a position of certainty. Hence, whilst there is never any room for claims to authority, it is the case that reason enables us to reveal our mistakes and take a step nearer to the truth we are seeking to establish. We can use rational criticism "to contemplate and adjudicate, and to discriminate between, competing theories" (p.384). By using such standards we begin to realise how little we know, and we learn to look critically at our own thinking.

Two other pieces of research which have been published more recently have made a determined attempt to establish some kind of rational basis upon which we might be able to escape from this increasing sense that knowledge is essentially both subjective and relative. The first of these was written by an anthropologist (Brown 1991), specifically as a challenge to the dominant paradigm of cultural relativism. Brown's view was that the social sciences had buried the idea of universals at the end of the nineteenth century in favour of cultural relativism and that the time had come to re-examine the issue, and the evidence, in a kind of Kuhnian paradigm-shift. Brown argues that "it is wrong to think that there is some kind of zero-sum game – or even worse, a winner-takes-all game – between universals and the culturally particular" (p.156). Brown suggests that we are so imprisoned in our devotion to postmodernist relativism that we have, perhaps almost unconsciously, rejected the possibility that there might at least be *some* universals which unite us across diverse cultures, perspectives, ideologies and value-systems, and he cites a considerable amount of research evidence to support this assertion. If Brown is right, then we might have within our

grasp a basis for at least some kind of shared meanings, and therefore the possibility of some kind of grounds upon which we might begin to establish a consensus about what constitutes valid knowledge about the world.

The second piece of research comes from the world of psychometrics. Psychometrics has had to face increasingly serious attacks upon its credibility in the last fifty years or so. This is partly because it tends to originate from positivist assumptions about the world, and correspondence theories of truth, but more importantly because it reduces a whole range of human behaviours to a series of quantitative measurements. This is particularly problematic in education. Kline (1998) argues that "claims by social scientists that there is no reality, that reality is a social construct, are unfounded and confused" (p.22). Kline's argument is that poor measurement methods have been the root of the problem for psychometrics in the past but that a whole new battery of statistical testing methods, such as factor analysis and multi-level modelling, enable us to take account of multiple variables that, in the past, we have been unable to control. New technology has also provided us with the kind of computing power which makes this kind of analysis possible. Kline claims that by using these methods it is possible to identify "a small number of factors of psychological significance" (p.179), such as fluid and crystallised intelligence, and that therefore "it still makes sense to think of some kind of external reality and thus to hold to a correspondence theory of truth" (p.179).

WHAT CAN BE KNOWN? EVIDENCE FROM POSTMODERNISM

Postmodernism is not easy to define. In essence it rejects what came to be known as "modern" during the twentieth century and accepts that we now live in a culture of endlessly contested meanings, contradictions, paradoxes, and even a kind of "hyperreality". The term postmodern is used to group together ideas from a very wide variety of sources but all of them,

essentially, claim to reject the traditional grand-narratives that have been used to investigate and explain our experience of the world in the past.

Barthes (1974), for instance, seriously undermined the idea that a written text might be capable of embodying some kind of definitive statements with regard to our knowledge of the world. He suggested that, in effect, once any text has been written, and released into the world, it takes on a life of its own. A host of different interpretations of the text are made by different readers, and commentators, so that whatever meaning the author may have originally intended quickly becomes obscured and even open to question. The reader, in effect, always brings something to the text which influences the way that she/he interprets that text. It is interesting to note that we see this process in operation even where sacred religious texts are concerned where, despite the beliefs of some followers that the text represents some kind of revealed truth from beyond normal human experience, there seem to be few, if any, limits on the variety of interpretations that any population of believers will lay claim to.

Barthes (1993) also introduced the idea of semiotics, which he believed give us the capacity to "decode" culture around us. Semiotics is concerned with the information inherent in signs, symbols, gestures, and so on, which give us hidden clues about the world, which we can learn to read. Barthes was concerned about the way in which "newspapers, art and common-sense constantly dress up a reality" (p.11) and he claimed that he wanted to "track down, in the decorative display of *what-goes-without-saying*, the ideological abuse which, in my view, is hidden there" (p.11).

For Lyotard (1984) postmodernism is "incredulity towards meta-narratives" which are used to give legitimacy to certain types of knowledge, most notably science, which legitimates itself by asserting that it is open to challenge and refutation. Given that many of the meta-narratives that legitimated various types of knowledge in the past no longer have any

authority, one obvious question is where legitimacy now comes from. We find ourselves in a society, according to Lyotard, where we all live at the intersection of many of these narratives. Knowledge has become a commodity, which has exchange-value, in the same way that other products have exchange-value, and therefore "knowledge and power are simply two sides of the same coin" (p. 9). We therefore need to be aware that the ruling class in our society "is the class of decision-makers" (p.14). Our "knowledge" is fundamentally based in the kind of language games which Wittgenstein suggested that we play. Our social experience is "a flexible network of language games" (p.17) in which the "moves" that we make, and others make, just like the moves we have in games, are all constantly inter-acting and producing counter-moves, so that we are all "located at 'nodal points' of specific communication circuits" (p.15). The new legitimacy for knowledge therefore becomes consensus. Through deliberation we decide what are the norms by which any type of knowledge will be regarded as legitimate or not. Performativity, in a complex technological society is a language game in which the "goal is no longer truth" (p.46) but "the best possible input-output equation" (p.46). The state can no longer legitimise knowledge using the narratives of the past so, in order to maintain power, "scientists, technicians and instruments are purchased, not to find truth, but to augment power" (p. 46). As our technological efficiency becomes ever more advanced, especially with computer technology, "performativity increases the ability to produce proof" and to "be right" and this "cannot fail to influence the truth criterion" (p.46).

Foucault (1980) sees truth as "linked in a circular relation with systems of power which produce and sustain it, and to effects of power which it induces and which extend it" (p.133). Power and knowledge are inextricably linked together. This produces a "regime of truth" (p.133) and the job of the intellectual is to constitute "a new politics of truth" (p.133) in which we detach "the power of truth from the forms of hegemony, social, economic and cultural, within which it operates at the present time" (p.133). For

Foucault the fundamental problem in our society is not "error, illusion, alienated consciousness or ideology; it is truth itself" (p.133). The battle is about the "status of truth and the economic and political role it plays" (p.132).

More radical amongst postmodernists are Derrida and Baudrillard. Derrida (1978) questions the very basis upon which our intellectual systems are built: reason itself. Logocentrism is the method by which we use reason to try to represent the world. The idea that reason can lead us to "truth" is anathema to Derrida and we need to deconstruct the methods by which "truth" is created. The assumption that language has stable meanings can no longer be accepted. By the playful use of language we can constantly recreate meaning. The problem for Derrida is that he uses reasoned argument to attack the very idea of reason itself. This would appear to be fundamentally irrational, leaving us with no tools with which to investigate, or think about, the world. If knowledge is simply a perpetual game of re-invention what basis can there be for making decisions?

For Baudrillard (1983) we live in a "hyperreal" world of "simulacra" (p.2), which develop through four successive phases: a reflection of a basic reality, a masking and perverting of a basic reality, a masking of the absence of any reality, and finally a pure simulacrum which "bears no relation to any reality whatever" (p.11). Disneyland, for Baudrillard is a "perfect model of all the entangled orders of simulation" (p.23). It is an image of "real" America which bears no relation to reality itself. Our increasing capacity to produce virtual realities means that "the hyper-realism of simulation is expressed everywhere by the real's striking resemblance to itself" (p.45). It may well be that our lived experience at the beginning of the twenty-first century lends support to these ideas. Perhaps we are beginning to feel, at the very least, that the boundary between reality and unreality is becoming increasingly blurred. If Baudrillard is right that hyperreality is replacing any kind of reality that we ever believed in, then we are essentially living in a

world of increasing illusions by most traditional epistemological standards. As we will see, the concept of targets in education could be regarded as an example of just this kind of hyper-reality.

ESTABLISHING AN EDUCATIONAL RESEARCH RATIONALE

In the nineteenth century, and the first half of the twentieth century, educational research was dominated to a large extent by the positivist paradigm with its attendant ideas that there is a reality which can be objectively measured, and that laws and predictions can be formulated as part of this process of measurement. Nowhere was this more evident than in the study of intelligence. Black (2001), for instance, shows how intelligence tests originally developed for modest diagnostic purposes rapidly became an instrument of measurement for entire populations, and that all sorts of inferences were then made from such measurements, not just about the future potential of individuals, but even more seriously, about the relative value of individuals in society. However, since coaching can improve scores on these types of tests, and there are significant problems with bias, especially cultural bias, in such tests, questions began to be raised about what intelligence actually is. If we are unable to define, with any degree of confidence or accuracy, what intelligence is, how can we measure it? Perhaps intelligence is nothing more than the ability to do well on intelligence tests. Gardner (1993) has even suggested that there are many types of intelligence, physical, creative, and so on. Goleman (1996) argues that our view of human intelligence is far too narrow and that our emotions play a far more important part in our thought than we have been prepared to acknowledge. Those with "emotional intelligence", he argues, excel far more than those with traditional forms of intelligence as measured by IQ tests. Furthermore, Goleman argues that this can be learnt, and that the very architecture of the brain is changed by the way that children are brought up. This clearly raises fundamental questions about models of ability based on positivist assumptions.

A second crisis for positivist assumptions in educational research can be seen in the behaviourist influences on education. Having established his "schedules of reinforcement" for behaviour, Skinner (1968) published a book called *The Technology of Teaching* in which, he described a series of "teaching machines" constructed using the principles of operant conditioning. The assumption was that there are too many intervening variables in the traditional classroom, and that conditioning techniques require scientific control of as many variables as possible, if successful learning is to take place. Skinner opined that "educators are seldom willing to concede that they are engaged in the control of human behaviour" (p.259) because they suffer from "a general fear of power" (p.259). The implications of these kinds of positivist and reductionist assumptions hardly need rehearsing and it is little surprise, therefore, that a reaction against the positivist paradigm set in during the second half of the twentieth century.

The ideas which underpinned the interpretive research model, which increasingly displaced positivist approaches, are clearly embodied in *Knowledge and Control* by Young (1971). The aim was to move away from positivist assumptions about establishing the nature of some external reality by objective investigation, and move instead to examine the experience, beliefs and perceptions of Plato's prisoners chained facing the cave wall. It was assumed that exploration of the subjective experience of actors in any social situation would uncover the ways in which they made sense of their own experience, and that this would reveal what motives and intentions lay behind actions. So, according to Carr and Kemmis (1986), the "task of 'interpretive' social science is to discover these meanings and so make action intelligible" (p.88). The intended outcome of such research is to illuminate, to enlighten, and understand better what is going on and why.

On the face of it, this appears to be a much more credible research approach, not least because it rejects the idea that human capacities, or experience, can

be easily measured or understood, or that outcomes can be easily predicted. However, interpretive approaches also have considerable problems inherent within them. It is difficult to produce useful generalisations, for instance, because findings are so situation-specific. It is also difficult to provide objective standards by which such research claims can be evaluated. Unlike positivist research, where replication often makes it possible to check findings, in interpretive research this is effectively out of the question. Questions arise about potential self-deception by participants, about how truthful participants are likely to be, about the effects of the researcher on participants, about demand characteristics, about the agendas created by the researcher and also about the kinds of assumptions which are being made by both researcher and participant, both consciously and unconsciously, about the issue under consideration. The dilemma is best understood by reference to an analogy from quantum physics noted earlier: the very act of intervening in a sub-atomic situation, in order to study it, changes the behaviour of the particles being studied. The same problem exists for the interpretive researcher and it is difficult, if not impossible, to ever really quantify these effects.

Where does this leave us in terms of trying to establish an educational research rationale? One attempt to reconcile these conflicting research methodologies has been made by Habermas (1987) whose critical social science aims to harness the strengths of both methodologies. The intention is to provide actors in social situations with "emancipatory knowledge" about their situation, knowledge which would also provide them with the kinds of insights needed to take action to change situations. Unlike interpretive approaches which merely aimed to "enlighten" social actors, critical theory aims to prepare human beings to implement change. Similarly, unlike positivist approaches, the aim is not to describe an external reality, but to show how such conceptions impinge on individuals and affect the way that they see themselves and the world. An essential part of this

process is to provide an "ideal speech situation" where the absence of coercion and irrationality allows a rational consensus to emerge.

Whilst this may appeal to our sense of democracy, equality and freedom of speech, it must be open to question as to whether it is possible to create any situation in which people would feel completely unconstrained about expressing their views and ideas. Furthermore we might view the idea that a consensus is possible as being somewhat idealistic, especially given the kind of highly contested value-perspectives which education finds itself in the midst of. Habermas seems to believe that a consensus will emerge: experience in educational debate suggests that this is far from certain.

RATIONALE FOR THIS STUDY AND RESEARCH METHOD ADOPTED

Whilst there are many problems surrounding the use of both positivist and interpretive research methods, it could be argued that they may be complementary if used together, in that each approach may be able to reveal a range of different insights and perspectives. However, since both of these methods embody a whole range of potential assumptions and values, it is also clear that evaluating the findings from either method involves a willingness to be as rigorous, as open, and as honest as it is possible to be, about the ways in which these values and assumptions colour any interpretations of the research findings. However hard we may try to take up a value-neutral position in evaluating research findings, we have to accept, as Hammersley (1999) notes, that claims to value-neutrality often mask underlying value-commitments. Hammersley does not believe that we should therefore abandon any commitment to value-neutrality, and simply write from an openly ideological position, but rather that we should strive to work within the *ideal* of value-neutrality because it "provides the basis for collective assessment of the possibility of bias in the work of researchers, on the part of their colleagues and of others" (p.27). For Hammersley, research

is fundamentally bound up with the idea that it is possible to produce some kind of universally valid knowledge if a commitment to value-neutrality underpins the research, however difficult this may be to achieve, and however inadequate attempts to do so may appear to be. Walsh (1999) takes a similar line, suggesting that we need to self-critically engage with our own ideological positions and that, whilst doing this, we cannot guarantee any kind of value-freedom, we can still retain a commitment to the *idea* of objectivity. This represents the position adopted in this research, together with Popper's concept of knowledge as always being provisional. With regard to psychometric knowledge claims (such as those made for the CAT), the assumption is that Kline (1998) is correct in arguing that whilst psychometric grand-narratives may be seriously open to question, there are certain knowledge claims which are demonstrably sustainable.

In terms of both method of inquiry, and methods of evaluation, the approach used in this study has many foundations in the work of Guba and Lincoln (1989) where a form of "hermeneutic dialogue" (p.17) has been employed which assumes that all the stakeholders involved in the situation being researched have viewpoints which represent "meaningful constructions that individual actors or groups of actors form to 'make sense' of the situations in which they find themselves" (p.8). Since contexts affect individual constructions of the world, and are affected by them, an attempt has been made to include a range of contextual information in the research report, both in Chapter 1, and in the analysis and discussion of the research findings. The research therefore attempts, as Guba and Lincoln recommend, to be more than just information gathering, in recognition of the fact that meanings are socially constructed and that meanings can only be understood within their social contexts. Evaluation of the research will also draw on Guba and Lincoln's criteria of credibility, transferability, dependability and confirmability. However, the view of Guba and Lincoln that knowledge is entirely socially constructed is problematic and the position adopted in this research is underpinned by Scott's (2000) concept of transcendental realism

where, despite changing epistemological conceptions, certain persistent ontological features almost certainly can be identified. Given the nature of education, and particularly the subject matter of this study, it is assumed that these features are more likely to have emergent properties than be directly observable.

Some further postmodernist ideas also underpin the rationale for this study. Usher (1999) underlines how postmodernism is prepared to accept that knowledge of the world is uncertain and ambiguous. Knowledge is neither neutral, or a mirror of reality, but is produced within an authorial context where the author recognises, or is at least prepared to examine and challenge, his/her own assumptions about how things are, and about how they should be researched. One important element in enabling others to evaluate research might well be autobiographical information about the author and this is the reason why this information has been included in Chapter 1. Scott (1999) argues that we need to reflexively understand where we are positioned as researchers. If we reject correspondence theories of truth then "we are forced to confront the embeddedness of our ways of knowing within social and political arrangements" (p.121). The idea in postmodernism is that no discourse can ever capture the whole truth about a situation, so the postmodernist foregrounds differences, and opens up perspectives, rather than attempting to set up alternative positions. This also represents an ideal to which this study has attempted to remain faithful.

Given Lyotard's suggestion that we all live at various intersections of a whole range of different narratives, this research started out from the assumption that a wide range of perspectives needed to be examined. Since the concepts of value-added, and target-setting have important implications for students, teachers, parents, and more widely for schools and society generally, methods were employed which opened up the possibility of being able to access, at least to some extent, the various narratives and narrative intersections in operation for these various groups. At a wider level, the

issue of government narratives is equally important, especially given Foucault's and Lyotard's concerns about the power/knowledge relationship, so a range of government documents and writings were examined in some detail.

METHODS

The three broad research questions, identified in Chapter 1, were:

1 What insights can be gained with regard to the validity of value-added and target-setting procedures at Key Stage 3?

2 What insights can be gained with regard to the reliability of value-added and target-setting procedures at Key Stage 3?

3 What are the philosophical and practical implications of these insights?

In order to investigate possible answers to these questions quantitative data were collected for the entire year-group population of students who entered the school between 1994 and 1998. These data sets included Key Stage 2 test results and teacher assessments, scores on the Cognitive Abilities Test (CAT), and Key Stage 3 test results and teacher assessments (including those for non-core subjects). For the intake years studied, therefore, the data collected related to tests and assessments completed between 1997 and 2001. Where data for individual students was incomplete, the entire data set for that student was rejected. (The size of the resultant population for each year group, and the ability profile, can be seen in Appendices 7-26, and population sizes are also quoted in Chapter 4).

As we will see in Chapter 3, successive government documents and reports have argued that linear regression is a valid and reliable method by which individual schools can calculate their own value-added, despite the evidence

to the contrary. The aim of the research was therefore to use just such a procedure, in one school, in order to ascertain how valid, and how reliable, such a procedure might be. Linear regression was therefore used to analyse the nature of the relationships between the various sub-sets of data collected, and to examine the stability of these relationships over the time period under investigation. The findings were then compared with national data in order to ascertain what further insights this might reveal in terms of the validity and reliability of the data, and linear regression, as the basis of value-added and target-setting procedures.

Further quantitative data were also collected in the form of questionnaires about targets and target-setting distributed to students at the beginning of Year 7 (soon after they had been given their targets) and at the end of Year 9 (after they had been given their Key Stage 3 results), during the academic years beginning in 1999, 2000 and 2001 (see Appendices 1 and 2 for copies of these questionnaires). These data were again collected from the entire year-group population present in school on the day of administration. (The total numbers of respondents can be found in Chapter 4). The questions were designed to collect data on a range of student perceptions with regard to targets, and the target-setting process, and to try to quantify the possible effects of targets on motivation and attitudes. Questionnaires were also distributed to staff in the summer of 2000, and parents in the summer of 2001, to gather further data on the perceptions and attitudes of these stakeholders in the process (see Appendices 3 and 4 for copies of these questionnaires). The resultant sample in these two cases was therefore self-selecting (see Chapter 4 for actual resultant sample sizes). The data from all of these questionnaires were analysed using percentages, in order to ascertain whether any particular patterns were apparent in the data sets, and estimates were made with respect to the validity and reliability of these patterns, given the various population and sample sizes.

Quantitative data was also collected with regard to actual student performance at Key Stage 3, in relation to their targets, for the entire populations for the academic years beginning in 1999, 2000 and 2001. Percentages were calculated with regard to the numbers of students achieving at, above, or below their targets, in both Key Stage 3 tests, and in the teacher assessed subjects, in order to see whether any patterns were apparent in the data.

Qualitative data was collected through a series of in-depth interviews with a group of Year 9 students in 2002. This was the first cohort of students to have been given their targets at the beginning of Year 7, and it was therefore possible to gather information about the longer terms effects of targets on students. Two students were interviewed from each of the four target levels used in the school (levels 4-7) at Key Stage 3 (see Chapter 1 for the ability levels these students represented). The students were selected by their own subject teachers on the basis of their likelihood of being able to effectively deal with the issues, and express their ideas. A semi-structured interview technique was used (see Appendix 5 for the outline structure of the questions used). Initial questions were formulated after analysis of the questionnaires used to collect quantitative data from students, staff and parents. Short-hand notes were taken during the interviews and transcribed immediately after the interviews had been completed. The use of taped recordings was considered but it seemed possible, from knowledge of the students, that this might inhibit some of their responses. (Tape-recording was not used in the staff interviews either, for very particular reasons, as we will see below and it was assumed that the validity of the procedures might be impaired if one set of interviews was taped and another set not taped). Questions aimed to open up a wide range of perceptions and attitudes with regard to the effects of targets including effects on motivation, perceptions, classroom relationships and parental attitudes. Supplementary questions were frequently used in an attempt to open up further discussion, but questions were always framed in such a way as to try to avoid, as far as

possible, any prompting, either explicit or implicit. The use of leading questions was studiously avoided. Detailed analysis was then undertaken in order to try to identify a range of themes through which the findings could be reported. Considerable care was taken to ensure that the report of the findings reflected the widest possible range of viewpoints.

Qualitative data was also collected through a series of interviews with a cross-section of eight staff covering all faculty areas within the school, and the full range of positions within the staffing structure of the school. The faculty areas covered were Modern Languages, English, Information Technology, Design Technology, Special Needs, Humanities, Mathematics and Science. The positions which these eight staff occupied within the school included a main grade teacher, a teacher with one responsibility point, a second in faculty, a head of department, a pastoral head, two faculty heads, and a senior teacher. Semi-structured interviews were used, as for the student interviews, (see Appendix 6 for the questions used to initiate discussions). As with the student interviews, initial questions were formulated after analysis of the questionnaires used to collect quantitative data from students, staff and parents. Short-hand notes were again taken, and transcribed immediately after the interview. Once again, taped recordings were considered, but initial soundings with staff suggested that they might be less willing to open up wider issues if recordings were used and the procedure was therefore rejected. (There was a particularly fraught atmosphere in the school at the time with regard to discussing such issues openly, and many staff felt unhappy about their views being recorded onto tape). As with the student interviews, once all the data had been collected, a number of themes were identified through which a report could be constructed which would reflect the widest possible range of responses expressed during the interviews.

SUMMARY

It was pointed out in Chapter 1 that the concept of educational assessment is both highly problematic, and highly contested. Since value-added and target-setting procedures are formulated using the data obtained in such assessments, it is not surprising to find that these procedures give rise to even more highly contested claims by the various stakeholders involved. The aim of Chapter 3 is therefore to examine the ways in which these claims are contested, both in terms of assessment generally, and more particularly in terms of the validity and reliability of value-added and target-setting procedures. These theoretical perspectives will then form the basis of a wider discussion of the results of this investigation in Chapter 6.

CHAPTER 3

THEORETICAL PERPSECTIVES

GENERAL BACKGROUND

Assessment is a concept surrounded with difficulties, and as the purposes of assessment have become ever more diverse, the problems embedded in these various applications have become ever more apparent. Assessment can be called upon to measure, to certify or qualify, to help a teacher to understand what is going on in the learning process, to judge individual or institutional standards of teaching, to compare individuals or institutions, and more recently to measure how much "value" has been "added" by a teacher or an institution, above and beyond what might normally be expected, as well as to predict future performance for the purposes of setting targets.

Assessment, historically, has been viewed as a form of measurement which can reveal objective information about an individual but as Goldstein and Lewis (1996) note, we should be sceptical about claims of objectivity in assessment. In using assessment to raise standards, measure abnormality, or to secure some kind of equity, they point out that assessment is inextricably caught up in the values of the education system and the society in which it operates. Broadfoot (1996) argues that whilst we should be using assessment to help students to improve their learning, what we actually do is to largely use it for the purposes it was used for in the nineteenth century, namely to control and select. This conception of assessment, as we will see, has considerable implications both for the learner, and for the teaching and learning process itself. The tendency is that as we *increase* the emphasis on assessment as a form of measurement and accountability, we *decrease* the emphasis on assessment as a technique for diagnosing problems and actually improving the quality of learning and achievement. This conflict

between the so-called "summative" use of assessment as a measurement tool, and the "formative" use of assessment as a diagnostic tool, is hotly contested in education (Wiliam 2001). The debate has, at its heart, two issues fundamental to assessment, whatever shape or form it takes: validity and reliability.

VALIDITY

According to Messick (1993), who has written in considerable depth about validity, it is "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment" (p.13). It summarises the evidence for, and the potential consequences of, score interpretation and use. At the heart of the validity issue is therefore the question of the *inferences* we make from assessment data. It therefore involves a *judgement*. It is not an all-or-nothing concept. Any form of assessment can have a higher or lower level of validity depending on a number of factors. Inferences, according to Messick, are always *hypotheses*, and like any hypotheses, they need to be tested. However, whilst part of this testing will inevitably be experimental and statistical, given the nature of educational assessment, it will also be philosophical and, increasingly, political as assessment comes to be more centrally dominated and controlled from within the political process.

Messick identifies three traditional categories of validity. Content validity is concerned with the issue of how representative any test item is of the domain of knowledge which it claims to be measuring. Criterion-related validity is concerned with the relationship between test scores and some kind of criteria that we have in a specific applied setting. An example would be the relationship between scores in a final examination for medicine and a set of criteria for an effective practitioner in the field. There are, therefore, many types of criterion-related validities depending on the situation, and the

criteria, we are trying to make some judgements about. Construct validity is concerned with the degree of fit between information provided by a test, and the underlying theoretical rationale. A typical example would be the extent to which scores on an intelligence test can tell us anything about intelligence. If we believe that we can formulate a rational and coherent definition of intelligence, then we can make some judgement about the extent to which a test tells us something about how much of this trait an individual has. If there are problems surrounding the definition of such a construct we clearly face considerable problems in making inferences from scores on intelligence tests. Our conception of intelligence can easily be reduced to nothing more than the ability to do well on intelligence tests. If we believe that there are aspects of intelligence which such tests do not measure, such as, for instance, the concept of "emotional intelligence" (Goleman 1996), then making inferences from such tests becomes very problematic. Since many other constructs within education are just as highly contested, the validity of many procedures and inferences being made is, at the very least, open to question, and this should cause us to take considerable care about the assertions which we make, based on test scores. There are fundamental construct-validity problems surrounding the idea of value-added, let alone the concept of predicting future performance, and setting targets, on the basis of such evidence.

Messick believes that the social consequences of test use and interpretation have been neglected in traditional discussions about validity but that "evaluation of the consequences and side effects of testing is a key aspect of the validation of test use" (p.85). Put simply, any form of assessment and testing provides very strong messages, to those preparing to take the tests, about what is important and what is unimportant, and this inevitably affects the way in which students are taught, the way in which they learn, and the value which they put on the results. As we will see, the increasing emphasis on summative forms of assessment in many countries, is beginning to seriously distort the teaching and learning process. There are always

"backwash" effects from any assessment procedure (Gipps 1994), particularly once it has been in place for a period of time and the various actors in the situation have had a chance to determine the nature of the game that is being played out. Looking for patterns of examination questions, concentrating on some parts of the syllabus rather than others because they can potentially earn more marks, and learning particular techniques for success in a test or examination are all typical examples of these kinds of backwash effects with which we are all familiar. These kinds of feedback-loops inevitably begin to affect the construct-validity of any assessment procedure because, over a period of time, the assessment may well be more to do with measuring the capacity of an individual to learn the rules of the game than it is to do with measuring a particular domain of knowledge or skills. Furthermore, the feedback that occurs within any system of assessment or testing has important consequences in terms of the confidence and self-concept of any individual going through such procedures and this, in turn, is likely to affect future performance. Again, as we will see, there is evidence that subjecting students to national testing at the ages of seven, eleven and fourteen, not to mention sixteen, seventeen and eighteen, is producing precisely this kind of effect and it is therefore incumbent upon us to ask ourselves to what extent any given set of assessment procedures ultimately becomes a set of self-fulfilling prophecies. Whatever the answers to these very difficult questions, Messick is surely right in his assertion that "interpretations and uses of tests needs to be justified not only in scientific and professional terms, but in the public arena as well" (p.91). Unless these justifications are "communicated sensibly to non-professional participants in the public debate" we are likely to see an increasing use of the courts to settle such matters. As Linn (1993) notes, the legal system is increasingly being used to try to resolve many of these issues in the USA, and given the problems which the consequences of test use pose for assessment, it is likely, in his view, that such legal challenges will increase in the coming decades.

Another issue which is central to any discussion of validity is the question of the extent to which we can generalise performance on a test to other areas. For Messick (1993) "generalisability of the construct meaning of test scores across various contexts cannot be taken for granted" (p.56) and there is therefore a need for "systematic appraisal of context effects" (p.56), especially the degree of generalisability across groups, time-periods and domains of knowledge. Messick identifies four types of generalisation which we typically make. We can generalise from one population to another, from one setting to another, from one time to another, or from one task to another. Each of these types of generalisation carries particular validity risks and the issue which we need to examine is the extent to which there is a match between the situation we are generalising *from* and the situation we are generalising *to*. This has fundamental implications, as we will see, for value-added assumptions and procedures. For Messick this suggests that "studies of the transportability of measures and findings from one context to another should focus on identifying all the boundary variables that are a source of critical difference between the two contexts" (p.58).

For Crooks et al (1996) assessment can be seen as involving eight distinct stages, linked like a chain, and weakness in any of the links in this chain threatens the validity of the entire procedure. The first of these links is the administration of the test itself. There are clearly issues here of individual motivation, anxiety levels, how the tasks are communicated, and the conditions under which the assessment takes place. The second link is to do with the scoring of the test. Issues here might relate to whether the scoring is too analytic or too holistic, how the various forms and styles of response are treated, how consistent markers are, either as individuals, or as a group, and whether all marking criteria are applied consistently, or whether some are given more weight than others. The third link is to do with aggregation. Any test clearly involves the aggregation of marks across a series of questions, and different weights can be given to different areas of performance. Should

it be necessary, for instance, to achieve a pass mark on all tasks to pass an assessment, or is it acceptable to achieve highly in some areas, but not in others and pass on the basis of an average mark? Is it possible to aggregate marks on a diverse series of tasks? The fourth link is generalisation. To what extent can we generalise performance on any kind of assessment to other situations? Conditions of assessment are, after all, extremely variable, and given inconsistency in scoring, criteria, and the number of tasks which a procedure might involve, there must be a number of questions which we need to address if we are intending to generalise test results. The fifth link is extrapolation. The issue here is the extent to which we can assume that any given level of performance on a test represents competence within the area being assessed. Given that parts of the target domain will not be assessed, and given that the areas which are assessed will always be given various different weightings, the issue of extrapolation becomes a difficult one to resolve. The sixth link is evaluation. This concerns the extent to which a test shows a grasp of assessment concepts and their limitations. Inadequately supported construct interpretation, and biased interpretations or explanations of results, all threaten the validity of evaluative procedures. The seventh link is to do with decision making. Decisions made on the basis of inappropriate standards, or poor pedagogical decisions, represent further threats to the validity of the procedures involved. The last link is to do with the impact of the assessment. If positive consequences are not achieved, or if, even more seriously, negative impact occurs as a result of an assessment procedure, then once again validity is threatened. Teaching to the test is, according to Crooks, an example of impact validity. As we will see later, there are also much broader aspects of impact validity related to issues of the social effects of assessment, international comparisons, and accountability.

Crooks et al believe that weakness in any one of these links threatens the validity of the procedure overall. However, it is also the case that the relative importance that we give to each of these links will depend on how the assessment is to be used. For formative purposes, aggregation,

generalisation and extrapolation may be less important, but for summative purposes these same three links are likely to be extremely important. However, whilst different assessment purposes imply different validity emphases, Crooks et al are of the opinion that the administrative and impact links are often ignored in any discussion of the valid use of assessments.

There are a number of implications that arise from this discussion of validity. Theoretical and value-assumptions are clearly at the heart of validity. Even more fundamentally, according to Messick, "ideologies influence theoretical conceptions and test interpretation in subtle and not so subtle ways" (p.62) and we therefore "need to explore some ways and means of uncovering tacit value premisses and of coping with their consequences for test validation" (p.62). Since scientific observations are theory-laden and theories are value-laden, "both the objectivity of sensory experience and the logical distinctiveness of fact and value are undermined" (p.62). We therefore cannot afford to ignore the value implications which underpin any form of assessment when we are appraising the validity of score interpretations. We not only need to raise awareness of the consequences of these value-laden assumptions which underpin assessment, but the inferences which we make need to be "supported empirically and justified rationally" (p.63). Using test scores that appear to "work", without understanding what they mean, "is like using a drug that works without knowing its properties and reactions" (p.62). The likely outcome is that sooner or later we find ourselves having to deal with unexpected side-effects which could be quite serious. The social consequences of testing are a typical example. Validity therefore needs to be viewed as a unitary concept in which "appropriateness, meaningfulness, and usefulness of score-based inferences are inseparable and that the unifying force is empirically grounded in construct interpretation" (p.64). Of particular importance for value-added and target-setting procedures is the validity of predictive procedures. "There is simply no good way to judge the appropriateness, relevance, and usefulness of predictive inferences in the absence of evidence

as to what the predictor and criterion scores mean" (p.64). Value-added and target-setting procedures usually involve making inferences about new populations, based on inferences about existing populations but this, according to Messick, is "fraught with interpretive difficulties" (p.72). Motivation, age, and experience, to name just three factors amongst many, all contaminate the relationship between the performance of two populations.

RELIABILITY

Reliability is a far easier concept to deal with than validity. Whereas validity involves *judgement*, reliability is centrally concerned with the *consistency* of a measurement tool over a period of time, and it is therefore a much more mathematically-based concept. Since human performance inevitably varies even within a single test session, let alone over longer periods of time, inconsistency, arising from a very large number of factors, is endemic to any form of assessment. Reliability, according to Feldt and Brennan (1993), is therefore concerned with the "quantification of the consistency and inconsistency in examinee performance" (p.105). All measurement procedures must therefore be assumed to contain error and this error may be systematic, random, or even unpredictable. The standard error of measurement relates to inconsistency in the performance of a given individual. Reliability coefficients summarise inconsistency across several error-prone measurements. Coefficients lower than 0.70, according to Feldt and Brennan are generally regarded as not being suitable for evaluating educational performance. So-called "true scores" remain constant over time. However since the measurement process itself changes the examinee, as we have seen, "in the behavioural sciences the notion of a constant true score is frequently inconsistent with the manner in which measurement procedures are carried out" (p.107). Error variance can occur as a result of "random variation within each individual in health, motivation, mental efficiency, concentration, forgetfulness, carelessness, subjectivity or impulsiveness in

response" (p.107) as well as sheer luck as a result of random guessing. Variation can be from minute to minute, as well as over much longer periods of time. Situational, psychological and physical factors can affect reliability and in school testing situations "these effects can be significant" (p.107). In general, therefore, "reported reliability coefficients tend to overestimate the trustworthiness of educational measures" (p.108).

Error variance, according to Feldt and Brennan, is different at different score levels and should be reported. According to them raw score error variance is greatest in the middle of the score range and declines rapidly at the extremes of the score range. When raw scores undergo "non-linear transformations to various types of derived or norm-based scores, the trends might be markedly altered" (p.127). This has crucial implications for value-added procedures which, as we will see, frequently transform data in a variety of ways which raises quite serious questions about the reliability of the procedures involved. Value-added, of course, also involves a number of generalisability procedures and we therefore need to be clear about the relationship between the score variance in a test and the expected score variance in the universe to which we are trying to generalise. Given that value-added and target-setting procedures involve generalising over quite long time periods (typically two or three years) this presents formidable reliability problems.

VALIDITY AND RELIABILITY ISSUES IN NATIONAL CURRICULUM ASSESSMENT

National Curriculum assessment was conceived as a way of establishing valid and reliable methods by which local and national comparisons could be made of the performance of individual schools and local education authorities. It was also assumed that such forms of assessment would enable teachers and parents to evaluate the progress of their students and offspring against national standards. The problem with these assumptions, as

Cresswell (1996) points out, is that setting standards is always a process involving value-judgements and these values "influence every step of the assessment process" (p.57). These values can be seen to influence two fundamental decisions: *what* is to be assessed, and judgements about the *quality* of the responses. Cresswell takes the view that at least one of these values will change over a period of time and that therefore valid comparisons over time, (which is clearly at the heart of National Curriculum assessment), are impossible. Furthermore, what we measure, and the quality of the measurement "interact with each other at a deep level" (p.59) and therefore "absolute examination standards are seen to be a chimera; standards can only be defined for a particular purpose and on the basis of the shared values of a particular community at a particular time" (p.59). If this is the case, then comparisons become extremely problematic, especially over time. Indeed, Cresswell thinks that such comparisons become meaningless. Gray (1996), takes a similar view, commenting that there are few circumstances in which "the kinds of information researchers would ideally like to have available for the purpose of conducting comparisons can be obtained" (p.129). In Gray's view multi-level statistical modelling tends to "discredit the idea that finely ranked 'league tables' of schools' effectiveness are possible" (p.132) and that any individual school's effectiveness "is not a precisely estimable quantity" (p.132). When confidence levels are used in statistical analysis of data, this tends to show that the performance of some two-thirds to three-quarters of students cannot properly be distinguished from each other.

Woodhouse and Goldstein (1996) raise further questions about the issue of comparing institutions. In their view, comparisons based on aggregated data are very unstable because institutions are complex and constantly changing. If the models we use to make such comparisons omit any important factor, or even if the model is just inaccurate, then the validity of the comparison is seriously compromised. A typical example of inaccurate modelling, according to Woodhouse and Goldstein, would be the use of linear

statistical models when in fact the relationship between two sets of data represents a curve. (As we will see this has very important implications for value-added assumptions and procedures). A further problem is that "institutions have been found to be differentially effective for different kinds of student: they are also differentially effective for different kinds of outcome" (p.143). This suggests that there are complex patterns of variation between institutions which cannot easily be modelled. Comparisons based solely on National Curriculum assessments therefore raise significant validity and reliability problems.

In a study of the validity, dependability and reliability of National Curriculum assessment, Wiliam (1993), as we have noted earlier, voiced grave concerns about these issues. Domain definitions are seen as a particular problem by Wiliam because the various statements of attainment associated with the ten levels in the National Curriculum are, like any criteria, open to interpretation. Qualifying statements such as "simple", "regularly" and "common" are typical examples. Any criterion therefore has "what might be termed 'plasticity'" (p.341) and this means that "we have to use norms, however implicitly, in determining the appropriate standards" (p.342). (Norm-referencing is founded on the concept that in any large population there are statistically predictable percentages of individuals who will achieve at any given level of performance, and awards of grades or levels are based on these statistical assumptions. Criterion-referencing was introduced in an attempt to establish clear descriptive criteria which would be used to decide what grade or level should be awarded to any individual). Morrison et al (1994) express similar concerns about the use of criterion-referencing. A further problem is whether it is safe to assume that any individual who has achieved at say National Curriculum level 4, is also competent at the three lower levels. Given the problems with the criteria, Wiliam's view is that this assumption would be "unwarranted" and "untenable" (p.344). In terms of the reliability of these kinds of criterion-referencing, "the true-score variance can often be quite small and the

distribution of errors can be unusual" (p.346). In terms of aggregation of results within specific areas, the inferences that we can safely make depend to quite a large extent on the way that various sets of data are aggregated. For instance, since only a proportion of the criteria have to be achieved by a student before they are awarded a National Curriculum level, this tends to weaken "the warrant that we have for an inference based on the result" (p.347). In strongly scalable domains, such as Mathematics, where it is necessary to achieve competence at lower levels, before progress can be made at higher levels, we have a fairly strong warrant for the inferences that we make. In weakly scalable domains the validity of these inferences is clearly considerably diminished.

Brown et al (1997), in a study of the 1994 and 1995 National Curriculum tests at the end of Key Stage 2 found significant teacher concerns about the validity of the tests. Concerns included problems over the lack of time to complete the tests and the consequent anxiety that this produced in students. The tests were seen to be extremely difficult for Level 3 students and these students found themselves taking tests, much of the content of which they did not understand. Equally as fundamental was a concern about the mismatch between what appeared in the tests and what children had experienced in the classroom. The majority of the teachers who took part in the study were unhappy about this mismatch. As a result of these concerns, both primary and secondary schools showed a "reluctance to attach significance to the results" (p.284). Overall, even in these first two years of National Curriculum testing, schools were concerned about the validity of the test results, the emerging pressures on teachers to teach to the test, and the inevitable distorting effects which this was having on the curriculum. As we will see, these problems are still with us.

Apart from English, Mathematics and Science, assessment in the other seven National Curriculum subjects has had to depend entirely on teacher assessments, made at the end of each Key Stage. QCA (1998)

commissioned a report on the consistency of these teacher assessments following the publication to schools of guidance materials originally devised by SCAA. This found that most teachers believed that they were not achieving consistency of level judgements across schools "because of the different interpretations teachers give to level descriptions" (p.7). A particular problem was assigning levels to borderline students, and even more problematic was comparing students across different Key Stages because, whilst the National Curriculum was conceived as a continuous scale from Level 1 to Level 10, there often appeared to be little equivalence between the same level descriptors at two different Key Stages. It appeared that some teachers were using an average measure of performance across the year whilst other teachers were looking at performance on key aspects of the subjects being assessed. This lack of consistency, according to the report was leading to a difference of one level for identical pupils. Since each Key Stage essentially measures ability at four levels this could be seen to represent a potential twenty-five per cent possibility of error in the teacher assessment of any given individual. This has considerable implications if we are trying to make any estimates of value-added in the seven non-core subjects. In general, teachers seemed to think that consistency might be possible if there was more time, and if there was a possibility of cross moderation between schools.

McDonald (2001), in a study of test anxiety in school children found that overall the "anxiety-performance association increases as children become older" (p.96) and that fear of tests and examinations is widespread and is becoming more apparent as time passes. This clearly has a detrimental effect on student performance. Higher anxiety levels were seen to be associated with lower test performance across a range of subjects, and a range of studies suggested that somewhere between 4% and 9% of test performance could be explained by anxiety. "Surveys show that anxiety and concern surrounding educational evaluation, and tests in particular, is considerable" (p.92). Test situations "are a considerable cause of concern

and anxiety to a significant proportion of school children" (p.93) whatever methodology is used to study it. As McDonald notes, most types of fear decrease with age: evaluative and academic fears appear to increase.

The effects of National Curriculum testing on students have received little attention from researchers. Reay and Wiliam (1999) looked at the effects of Key Stage 2 tests on primary students in the last term before they sat the tests. Teacher anxieties were seen to increase student fears about the tests. Students expressed concern about the narrow focus of the tests and about their capacity to "produce their best under strict (and unfamiliar) test conditions" (p.346). Students had a very real impression that poor performance on the tests meant poor life prospects. The implications of this for self-esteem at Key Stage 3 are obvious and this may be a significant factor in the so-called Key Stage 3 dip. Strong currents of fear and anxiety were found in the conversations of children, irrespective of ability level, and as different ability levels became apparent, there was evidence of violence and aggression towards students who were likely to do well on the tests. Again it is tempting to wonder what effects are carried over to Key Stage 3 if this kind of experience is typical for more able students at Key Stage 2. Overall, Reay and Wiliam found a growing climate of hostility as the tests came closer, with teachers under pressure to improve results "irrespective of the consequences for students' achievement in wider terms" (p.352).

ALTERNATIVE CONCEPTIONS OF ASSESSMENT

Validity and reliability are always in tension with each other in any form of assessment and these tensions always play themselves out in the construction of any assessment procedure. Traditionally the emphasis has been on reliability, with consequently lower levels of validity. Since the National Curriculum was conceived on the assumption that it should be possible to make comparisons over time, it was clearly important that such

forms of assessment should be consistent over time, otherwise the comparisons would be invalidated. This has important consequences for the validity of such assessments. As we have seen, so-called "high stakes" formal tests and examinations give rise to a number of unwanted effects which then distort performance in various ways. Anxiety is a good example. How representative the test is of what the student has learnt, or actually knows, is another example. The more consistent and standardised any assessment procedure becomes, the greater the limits become on what *can* be assessed and *how* it can be assessed. If validity is reduced as a result of higher demands on reliability, one of the questions that inevitably arises is to what extent we can then generalise the results. How representative is performance in a very artificial situation like a test or examination of potential performance in the real world? Will a graduate student, in medicine for instance, who has done well in written examinations, make a good doctor? Is this even any guarantee of competence?

One possible solution to this was put forward by Gipps (1994) who takes the view that fitness for purpose should be a central concern in assessment. This is traditionally compromised, according to Gipps, by the influence of psychometric concepts such as norm-referencing, standardisation, and reliability, which compromise the validity of the procedures being used. This in turn, distorts the teaching and learning process and the content of the curriculum. Gipps therefore argues for "authentic" or "performance" assessment which is an integral part of the teaching and learning process and which she believes will "harness a powerful tool for learning" (p.176). This approach would enable the education system to move away from the idea of using assessment primarily to *measure* students, (summative assessment), and use it rather more to diagnose what it is going wrong with their learning, and provide them with the tools and insights which will enable them to *make progress* (formative assessment). Authentic assessments, suggests Gipps, would enable us to establish procedures which have much higher levels of validity. A teacher, for instance, could question

a student during an assessment. This is clearly not possible in a formal test. This would enable the teacher to get a much clearer idea about what a student really knows, and how they are thinking (Wood 1998). What might be perceived as a lack of understanding in a written answer during a formal test, might be perceived very differently in the context of authentic assessment. A further advantage would be that classroom based assessment of this kind would be much more realistic than the highly artificial world of formal, written assessments, and the validity of inferences about the potential performance of individuals in the real world would therefore be enhanced.

Broadfoot (1996) similarly argues that we need to move away from nineteenth-century, norm-referenced, conceptions of assessment and utilise forms of assessment which develop motivation, self-concept, self-esteem and metacognitive skills (the ability of the learner to reflect on, and evaluate, their own learning and thinking). For Broadfoot learning is a co-operative venture between teacher and learner, and assessment should be a central part of this process. We need this not only to enable us to help individual students to learn more effectively, but also because we are moving, according to Broadfoot, into a kind of society where individuals will need to be much more ready and willing to learn throughout their lives, and we therefore need to lay these foundations in the years of formal schooling. Governments therefore need to recognise and promote forms of assessment which really *empower* learners. One way of doing this, according to Broadfoot (1998), is through the Record of Achievement. This is designed to raise student motivation by involving them in their own goal-setting and by including in the assessment a much wider range of evidence of achievement than has been the case in traditional forms of assessment. (One variation on this is so-called "portfolio assessment" which has been quite widely used in a range of vocational qualifications). This is seen as an important method for improving the validity of assessment and for laying the foundations of life-long learning. High-stakes assessment, on the scale

on which we see it today, militates against empowering the learner, according to Broadfoot, and we therefore need to challenge existing assumptions about assessment and replace this "categoric" discourse with an "emancipatory" discourse of assessment.

There are clearly some powerful, and potentially seductive, arguments in favour of moves towards more "authentic" forms of assessment. There are also some very powerful counter-arguments. The issue of reliability cannot be ignored in more contextual types of assessment. Whilst the desire for higher levels of validity may command our support, if this is at the expense of reliability a number of serious consequences follow. Since assessment involves judgements about individuals, often involving consequences for their future prospects in life, it is not surprising that we expect any form of assessment to show a demonstrable level of consistency in its judgements. We also expect equity in these judgements. Whilst it is clear, as we have seen with National Curriculum assessment, that existing high-stakes formal assessments are frequently far from perfect in terms of their validity or reliability, it is difficult to see how "authentic" forms of assessment would necessarily be any better. Koretz (1998), commenting on the widespread use of portfolio assessment in the USA, highlighted a number of problems with this type of assessment. For those marking portfolios, differentiation between different levels of achievement "was in effect random" (p.322) and there was no evidence that markers could "agree about the relative quality of pieces" (p.322). There was considerable unreliability as a result of the variety of tasks that were sampled, and even the validity of what was being assessed was in question because "in many instances the relationships shown by the scores offered no evidence of validity" (p.327). Given that these types of assessment were conceived as a way of improving validity this is an interesting finding. Koretz concluded that whilst portfolios may well be useful for internal forms of assessment, validity and reliability were too low to make them of any real use externally.

Cumming and Maxwell (1999), whilst accepting that "it is as difficult to be against authentic assessment as it is to be against apple pie and motherhood" (p.178) raise serious questions about how authentic this type of assessment actually is, in practice. They question how realistic it is to assume that it is possible to create real-world situations in classroom assessments, because in the real-world there are a multitude of pressures, influences, and considerations which impact on our performance, which we cannot recreate in the classroom. Complex tasks and skills are often broken down into component tasks and skills in the classroom and this rather undermines the concept of authentic assessment. Classroom assessments, according to Cumming and Maxwell, are often "camouflaged" (p.189) as real-world situations but this is "a form of deceit" (p.189). Many types of authentic assessment therefore "verge on labelling, and exultation by association, rather than enactment of fundamental principles" (p.192). Once again we see the claims for higher levels of validity open to challenge.

Wolf (1998) even questioned whether portfolio assessment improved the quality of teaching and learning, suggesting that students spend more time hunting for information to put in their portfolios than they do actually learning anything, and that coping and cheating strategies were increasingly apparent. This type of assessment produced an immense workload for the teachers involved, and reliability levels were so low that a third of the teachers thought that students who were awarded a pass should have failed. Nearly a third of teachers were even using standards to mark work which were not part of the formal criteria for assessment. External assessors were found to be so short of time that they were unable to check the standards of the internal assessments effectively, and the use of written criteria to judge work was fundamentally flawed because "one cannot, either in principle or in theory, develop written descriptions so tight that they can be applied reliably, by multiple assessors, to multiple assessment situations" (p.442).

Where does this leave us in terms of assessment choices if high-stakes formal assessments embody potentially serious validity problems, whilst authentic forms of assessment embody equally serious reliability problems? The answer, as so often, appears to be in maintaining a balance between the two. Validity and reliability are equally as important, and whilst it is unlikely that we will ever reach a situation where these two factors are balanced to the satisfaction of everybody, it is fairly clear that this is an ideal that we must continue to strive for.

The same tensions are played out in debates about the balance between summative and formative types of assessment. The arguments in favour of formative types of assessment are considerable, as Black and Wiliam (1998) have shown, in a wide-ranging review of several hundred studies of the subject world-wide. "Gains in achievement appear to be quite considerable" (p.61) and "amongst the largest ever reported for educational interventions" (p.61), but as Sebatane (1998) has pointed out, exhortations to use more formative types of assessment are rather easier than changing classroom practice, because effective formative assessment involves a radical change of attitudes and viewpoints for both teacher and student. Sadler (1998) echoes these concerns, pointing out how crucial the quality of interactions is between the teacher and learner when employing formative assessment techniques. It is difficult to either identify, or quantify, exactly what gains are made through formative assessment, and it could be argued that it contributes as much to the culture of performativity, which we will discuss later, as summative assessment does.

In general the literature on formative assessment takes the view that this type of assessment has been increasingly marginalised by the greater and greater demands for summative types of assessment in the second half of the twentieth century (Black 1986), and experience suggests that this remains a significant problem in the minds of many teachers. It is tempting to wonder, though, whether this is the whole story. Perrenoud (1991) points out that

without better theoretical models of how situated learning occurs it is difficult to specify what formative assessment is exactly, or what it should be. "Parts of the feedback given to pupils is like so many bottles thrown out to sea" (p.87) because in order to give effective feedback to students we need to find a window into the cognitive system of the learner, and understand how, and why, they are thinking the way that they are. This demands immense skill, insight and experience, not to mention empathy and even intuition.

VALUE-ADDED

Value-added is a concept which has been around for twenty years or more and was originally borrowed from economics. The context and rationale for the concept has changed dramatically since the 1980s as Saunders (1999) has noted, and is now widely used to mean a fairer or more accurate measurement of student performance, and therefore the quality of their education. The history of the concept, however, shows a range of quite diverse and conflicting meanings. Sometimes it is regarded as a tool for measuring *progress*, and at other times it is seen as a yardstick by which to make *judgements*. These ambiguities reflect its origins in economics. The underlying political agenda is concerned with raising educational standards, value-for-money for the taxpayer, and professional accountability, but as we will see later, in the discussion of research on accountability, procedures designed as a form of quality assurance can quickly become procedures of control, with important consequences in terms of professional autonomy and judgement. As Saunders points out, the globally competitive market-place in which we have found ourselves in recent years has produced a "consensus belief that education and training are important levers for economic competitive advantage" (p.234).

It is perhaps not surprising, therefore, that we see a model of education emerging in the 1990s which is founded on the industrial model, where the

difference between the quality of the *inputs* (raw materials) and the *outputs* (products) is what constitutes the *value-added*. This is a temptingly seductive concept. Take some raw timber, turn it into a piece of furniture, and you can make some estimate of the value that has been added (say by using the price at which it will sell in the market-place by comparison with other, similar products). The idea of value, however, is quite a slippery concept upon further examination. For example, the value we put on a piece of furniture may depend on quite a number of variables including our personal taste, what particular functions we want the piece of furniture to fulfil, whether we see it as a longer-term investment, and so on. The meaning of "value" in value-added is therefore quite problematic, even when we apply it to inanimate products like furniture. How much more problematic is it likely to be when we apply it to human beings? A factory has some quality control over its raw materials, and this has consequences in terms of the quality of the end product. State schools, generally, have little or no control over their "raw materials", their intake, which has similar consequences for the quality of their output. Factory raw materials are inanimate, malleable, and we can treat the materials in whatever way we choose, in order to produce the end product we want. Students have minds, wills, feelings, perceptions, memories, agendas and a whole host of other human attributes which makes them a very different kind of "raw material". Saunders therefore suggests that in education the "evaluative potential located within the term has been converted into an evaluative purpose" (p.237).

How do you measure value? We might evaluate our piece of furniture in aesthetic terms. Can we do this numerically? Perhaps, but we may well disagree about how to do this, and we would probably regard it as more effective to use descriptive terms. What can we measure numerically? Size, shape, weight, and price perhaps, but the things that *can* be measured numerically wouldn't tell us a great deal about the overall value of the product. If we were then to try to make comparisons between this product,

and a range of similar products, we would need to establish some common measures upon which to do this. What tends to happen, as this process becomes more complex, is that we tend to measure those things that can be easily measured, and therefore compared, and we tend to ignore those things that cannot be easily measured and compared, such as, for instance, comfort level. Because it is not easy to compare *judgements*, what Saunders suggests has happened in education is that we have "defaulted to using proxy variables which already exist and are measurable" (p.241). We therefore measure the qualifications of students upon entry to an institution, and we compare this with the qualifications upon leaving. By comparing one institution with another we can then begin to make some judgements about what value has been added by comparison with institutions with a similar intake ability. However, we might well ask whether the value that has been added is what the *student* has added, or what the *institution* has added.

Despite the inherent problems, value-added procedures did seem to represent a potential step towards a fairer way of comparing schools after the obvious inadequacies of league-tables were seen for what they were. The government commissioned a number of studies in the 1990s culminating in the *Value-added National Project*. SCAA (1997a) produced the final report on this project. Their definition of value-added was "the difference between a statistically predicted performance (based on prior attainment and the general pattern in the data) and the actual performance" (executive summary). This begs a number of questions. What is prior attainment and how is it measured? What are these general patterns in data? What are the statistical procedures involved, and what assumptions do they make? Typically, according to the report, there is a correlation of 0.7 between some prior attainment measure and a later measure of attainment (such as an examination result). This means that about 49% of the variation in examination results can be predicted. This seems to leave a considerable proportion of the variation which could be explained by a multitude of other factors to do with the school, social background, and the personality and

self-esteem of individual students, to mention only a few. The report accepts that the effectiveness of the system nationally will be "dependent upon confidence in the integrity of the data underpinning the system" (executive summary) and the way in which such data is interpreted. Herein lies what is, perhaps, the central difficulty with value-added procedures in education. How much confidence can we have in the integrity of the data, and how do we interpret the data when we have it?

If value-added procedures are to fulfil their purpose of greater equity and fairness in terms of comparing schools, the first thing that we need is valid and reliable *baseline data*. We need to measure all schools from the same starting point if the comparisons are to be valid. We also need to know that these measurements will remain stable over time if the comparisons are to be reliable. As we have already seen, the assumption that data with this level of integrity are available is somewhat open to question and yet the report by SCAA (1997a) states quite clearly that from 1998, Key Stage 3 outcomes should be predicted from Key Stage 2 test results. Yet elsewhere in the report we are told that "much of the variation in value-added measures will not be explicable in terms of any actions that the school has or has not taken" (p.68) and that "there may be pupils for whom the input measures were clearly inappropriate due to lack of effort, illness, or stress" (p.68). Issues of attendance, students who are geographically mobile in terms of their education, and issues of the stability of data over time are all examined by the report. Given the acknowledged range of factors which can therefore affect the integrity of the data, it is pertinent to ask how the report can conclude that a national system of value-added should be implemented when "if the input measure is in error the interpretation of the value-added is in question" (p.125).

Because of the perceived problems with regard to the validity and reliability of Key Stage test data, one measure which many secondary schools have attempted to use as baseline data is the Cognitive Abilities Test (CAT)

published by the NFER (1986). These are age-standardised tests, rather like the traditional intelligence tests, which use the same set of questions year on year, and therefore afford at least some possibility of using the results as a reliable baseline measure, both for comparing schools, and for comparing results across a period of time. The tests consist of a verbal reasoning battery, a numerical reasoning battery and what is called a "non verbal" reasoning battery (which involves completing tasks using shapes and space). The first two of these tests measure what is generally known as "crystallised intelligence", in other words intelligence developed as a result of formal schooling. The final test measures what is known as "fluid intelligence", in other words innate ability. NFER (1986) reports good reliability figures for these tests. Using a sample of 11,287 student results, the Kuder-Richardson reliability coefficients (using Formula Number 21) ranged from 0.90 to 0.97, depending on the age and level at which the various subtests were taken. It would appear, therefore, that this represents rather more reliable baseline data for value-added procedures, especially at Key Stage 3, than Key Stage 2 test data would. It is somewhat surprising, therefore to find SCAA (1997c) on the one hand accepting that schools in previous years have been very concerned about Key Stage 2 test data, and that "there has been some justification for these criticisms" (p.2), and yet on the other hand claiming that "Key Stage 2 tests are now sufficiently well established and reliable" (p.5) that we can have confidence in the results and that therefore "commercially produced standardised tests in Year 7" (p.5) are no longer necessary. The report accepts that a common criticism of Key Stage 2 tests is that "the levels are not equivalent to the same levels at Key Stage 3" (p.5) but we are assured that procedures are in place for bringing about these equivalences and that from 1998 reliable and valid national comparisons will be possible. As we will see, the results of this study, and the NFER analysis of CAT and Key Stage 3 data nationally, casts some doubt on these claims.

A further problem in using National Curriculum test data is that at each Key Stage students are generally awarded one of four levels. This represents a relatively crude measure for value-added procedures as SCAA (1994) admit: curriculum levels "should not be used in value-added studies since they are likely to lead to less reliable predictions and greater errors for individual schools" (p.11). What is needed according to SCAA is "finely differentiated measures of prior achievement" (p.20). Use of CAT data is one way of providing this kind of differentiated baseline data, but output data, equally, needs to be differentiated too. As we will see, although National Curriculum marks, as well as test levels, have been released to schools more recently, this seems to create as many problems as it solves, not least because of the different levels of entry in Mathematics and Science, with different mark schemes at each level of entry.

If we leave aside the question of baseline data, the next question we need to ask is about the statistical procedures which might be used to make a value-added calculation. The choice is essentially between what is known as linear regression and multi-level modelling. Linear regression correlates a measure of prior attainment, such as a test result, with a subsequent measure of attainment, such as another test result. In other words, it investigates the relationship between two variables. Multi-level models, by comparison, take into account a whole range of possible factors which might affect performance including gender, socio-economic status, whether or not a student receives free school meals, whether or not English is the first language of the student and so on. Clearly such models produce a much more complex picture of student performance and conclusions are far less easy to draw. Yet it would appear not unreasonable to assume that this kind of statistical procedure might well model what is going on with rather more sophistication than linear regression. Despite this, SCAA (1994) recommended that schools should use linear regression "using data aggregated to school level" (p.49). Yet in an appendix to the report it is admitted that "there is no single correct method of analysing a complex

social situation" (p.85) and that decisions about which variables to include in any analysis "depend on views about what is important in the real world situation we are attempting to model" (p.85). This potentially rather large defect in the recommended approach is dismissed with the observation that interpretation of the regression of multiple variables is highly contentious. Fitz-Gibbon (1996), in looking at performance indicators and value-added for schools accepts that schools consist of "self-organising units which thrive on the edge of chaos" (p.7) and that a large proportion of the variation in outcomes remains unexplained in a system as complex as education. Having accepted this reality Fitz-Gibbon then rejects multi-level modelling because it shrinks averages and this causes differences between institutions to disappear with populations of more than thirty. It is possible, however, that differences between institutions are not all that large once a wider number of variables have been taken into account. One has to ask how defensible it is to advocate a statistical model using only two variables, simply because this makes differences between schools clearer.

A study by the DFEE (1997b) adds further evidence to the argument against linear regression in value-added procedures. In examining the relationship between average GCSE points score and A level scores, for instance, the report concludes that the "relationship between the measures are different at the attainment extremes" (p.4) and that "a linear relationship may therefore not be a fair predictor of value-added for institutions with differing numbers of high and low achieving students" (p.4). Furthermore, different statistical models suggest widely different explanations about what does account for the variations in student results. The so-called intercept model suggests that 77% of the variation in A/AS level scores is due to differences between individual *students* within an institution and intake cohort, whilst only 22% of the variation can be attributed to differences between *institutions.* Use of linear and prior attainment models, on the other hand, suggests that 90% of the variation is between individual *students* and only 8% is due to differences between *institutions.* Gender differences are also a problem

because whilst there are clear differences, "the relationship between gender and progress is not linear" (p.7). Lower attaining boys and girls at GCSE make similar progress at A level, whilst higher achieving girls at GCSE make less progress than boys of similar ability at A level. The overall conclusion of the report was that "differential effectiveness judged against prior attainment is not consistent across institutions: the relationship between progress made by the brightest and the weakest students within the same institution varies a great deal" (p.21). This report concludes that "a multi-level approach is essential" (p.27).

Sammons et al (1997), in examining ideas about school effectiveness, also raise doubts about value-added procedures, suggesting that "the major difficulty in introducing a national framework for value-added measures is the lack of reliable standardised assessments to measure the prior attainment of students on entering school" (p.26), and that it may well be impossible to ever devise National Curriculum assessments which do meet this need, given all the other requirements they are expected to meet. What is needed, according to Sammons et al is a system of value-added calculation which separates and identifies "all the possible factors that may have an impact" (p.38) on attainment. They also draw attention to the varying levels of progress made by students at different ability levels, variations which remain concealed in many value-added models. There are a range of factors, as we have seen earlier, which may have some bearing on these variations, but Sammons et al point out that there are also *causal interactions* between all these factors, and many others, which we need to understand better if we really want to know what is going on. An input-process-output model, such as linear regression, based on measures of prior attainment, is therefore inadequate. Furthermore, longitudinal studies of schools are needed to back up this kind of investigation because effectiveness is a time-related, as well as an outcome-specific, concept.

Goldstein (2001) tested three different value-added models in a study of the performance of primary schools at Key Stages 1 and 2. This work suggested that very different inferences can be made about how much value a school is adding, depending on which model is used. Factors such as whether students are entitled to free school meals, for instance, when added into the calculations, produce very different outcomes from merely using test data from the two Key Stages and Goldstein concludes that value-added models "may lack information on further important variables" (p.439) and that "there is a great deal of inappropriate use of performance data for judging schools and teachers" (p.442). Goldstein and Sammons (1997), in a study of the effects of junior schooling on secondary school performance found significant effects, not just at Key Stage 3, but even at GCSE. When the junior school attended was added into the model it caused "the largest reduction in the secondary school variance" (p.228) and this suggests that junior schools "may be exerting a strong, persisting, influence" (p.228) on attainment at the age of sixteen. In their view, therefore, any value-added model failing to take account of the junior school attended "may be seriously deficient" (p.229).

As we have seen, an important issue that arises in connection with any form of educational measurement is how we use the data that it gives rise to, because this has important consequential effects for the validity of the procedure. In a study of the way in which schools use value-added data, Saunders (2000) raises some disturbing questions. "There are still many in the profession, at management as well as at classroom level, for whom such data present difficulties of acceptance, understanding and utilization" (p.245). Data was often summarised and incomplete in the way that it was presented to staff. At senior management levels, there was a frequent lack of reference to the data so that the conclusions drawn were often no more than "hunch or instinct" (p.249), and there was often a failure to appreciate that education is not always about "rational solutions to clearly defined problems" (p.249). At middle management level Saunders found a range of

approaches from deep analysis of the data, to try to improve the teaching and learning process, to those who simply saw it as a numbers game to be played. On the other hand, data were often regarded as a catalyst for discussion, rather than as some kind of literal truth. Value-added data were often seen as a threat, since it was seen as being part of the target-setting and accountability procedures, and staff often felt that they were cornered with the data, rather than it being used as the basis of a discussion and reflection on practice. Saunders suggests that we need to pay much more attention to the way in which value-added data are being used in schools, and that we take too much for granted.

This issue of the way that value-added data are used is also examined by Goldstein (2001). Value-added procedures, according to Goldstein, are based on the idea that assessments are objective and reliable enough to make valid comparisons. "Both these assumptions are questionable" (p.434). This claim for objectivity has often not been challenged, and it needs to be challenged because "these problems become acute when there are attempts to use test results to make comparisons over time" (p.434). Since there is no common yardstick over time, it is not possible to say whether change in a test score is due to change in performance, or in the test itself. Government publications make little effort to place such procedures in context "and give no hint that this is problematical" (p.435). The failure of the DFES to set out these limitations does not inspire confidence according to Goldstein. Since OFSTED use such data to inspect schools "it is not clear how such limited information can be used constructively whilst recognising its defects" (p.436). Goldstein goes on to raise serious concerns about the way that such data is now being used to appraise the performance of teachers as part of the new performance management procedures, particularly since this involves examining data over just a single year. To use such data to make judgements about the performance of individual students and teachers is extremely misleading because in secondary schools "it is very difficult to ascribe the progress of any one pupil in a given subject

to the teacher of that subject" (p.437). He concludes that current procedures need to be phased out because they "may be doing fundamental harm to education" (p.442).

TARGET-SETTING

The idea of national targets for education and training were originally promoted by the CBI in the early 1990s and this was subsequently taken up by the DFEE. According to the DFEE (1996) national targets and the National Curriculum "provide a context for target setting in all types of school (p.2). Target setting is defined in this report as "taking action by setting specific goals and targets assigned to raise educational standards" (p.5). Target setting, we are told, is "effective in schools which have taken a firm hold on school improvement" (p.5) because it "often places a greater onus on pupils to do better than on teachers to improve their effectiveness" (p.6). To be most effective, the report argues, teachers should be involved in setting the targets. However, lest target-setting should appear to be a cure-all, we are also told that "raising standards depends on a range of factors" (p.8), such as strong leadership and a constructive climate "and does not stem simply from the existence of target-setting strategies" (p.8). The key assumption is that setting targets causes teachers to focus more of their attention and effort on the quality of what they do in the classroom. This, as we will see, is at least open to question. There is a need to balance challenge and realism in setting targets, though quite how this balance is to be achieved is not clear. The report acknowledges that some head-teachers had already expressed concern at this time about targets being promoted in education "in connection with a need to improve Britain's industrial and commercial competitiveness" (p.35), but this concern is dismissed with the observation that "the desire to have all pupils achieve their full potential and the desire to improve the nation's competitiveness are not in conflict" (p.35). This would appear to many to be a quite extraordinary assertion

which is surely open to challenge at any number of levels. The idea that the good of the state and the good of individual are one and the same is actually more redolent of highly collectivist ideologies than it is of thinking in a modern liberal democracy. There may well be some overlap between the good of the individual and the good of the state, but to assert that they are one and the same carries some very disturbing overtones. The overall aim of the national targets, according to the report, is "to improve the UK's international competitiveness by raising standards and attainment levels in education and training to world class levels" (p.40). Educational targets are therefore not only inextricably linked with concepts of economic competitiveness, but they are actually *founded* on this concept. As we will see this has important consequences for the education system.

An early indication of the attitude of the then incoming Labour government can be seen in *From Targets to Action* (DFEE 1997a). This document reinforces the idea that performance targets can raise standards, and announces the intention of the new government to make targets compulsory in the forthcoming 1997 Education Act. Target-setting "leads to greater clarity and helps a school focus on pupil performance" (p.6). It can do this by setting reliability targets, which set a floor below which performance is not allowed to drop, by setting threshold targets, such as a percentage of students who will achieve a given level of performance, or by setting average targets. The report admits that each of these three different types of target tends to cause schools to focus on different types of students, and thus reveals an inherent problem in this apparently "simple and effective way of raising standards" (p.19). Further problems reveal themselves with the level at which the target should be set. Targets which are set too high, and which are not achieved, "can disappoint and undermine improvement initiatives" (p.16) so targets need to be challenging, but achievable. On the face of it this seems a fairly reasonable concept. In practice it is a lot more problematic. How do we know what is challenging, but achievable, especially when we are setting targets for performance at least two years

into the future, and three years into the future at Key Stage 3? The government answer is by using performance data, but already anticipating what this entails, the report immediately urges schools to "keep your system simple and avoid unnecessary bureaucracy" (p.23). A wise injunction, maybe, but how achievable is this? Targets, we are told, "should be based on sound evidence" (p.30). The purpose of the present study was to investigate just how "sound" this evidence might be. As we will see, the answer is far from convincing. Whatever the problems, however, this report is quite clear that "target-setting is an essential tool in school management" (p.30).

Given that the idea of educational targets has been with us now for at least six years, it is surprising that so little attention has been paid to the matter in the published literature. Flecknoe (2001) is clearly very disturbed by both the concept and the language surrounding it, suggesting that the concept of school improvement has now been narrowed down to nothing more than improving examination results. Many of the terms used in various government documents, according to Flecknoe, are vague, if not meaningless. The concept of targets which are demanding but achievable, for instance, tells teachers nothing about "how to set them and where to fix the standard" (p.218). Concepts such as "comfort zones" and "challenge zones" suggest that normative judgements have taken the place of "any reasoned investigation into the practice of target setting" (p.218). Underlying all this is an assumption that "greater information about everything that can be measured will lead to more effective measurement of learning and higher standards" (p.218). Unreasonable outcomes of these policies are ignored. Flecknoe argues that research suggests that what most people want from education is not measurable. People cite things like happiness, autonomy and morality and there seems to be a significant conflict between what governments seem to *assume* that people want from the education system and what people *actually* want. This leads the government to assume that "examination results must bear some

relationship to the outcomes that most parents require from schools" (p.220) but the idea that more, and better, examination results will produce a better society is, at the very least, open to question. In practical terms, there are real problems with target-setting for students. What is the effect on a student whose target is, say, a Grade F at GCSE? What does it feel like to be told this at the beginning of a two year course? Do we set the target higher to avoid demotivating the student? Flecknoe cites research which suggests that targets which are unachievable simply cause students to channel their energies into more anti-social kinds of achievement. If a target is going to be effective, therefore, it needs to be "valuable to the individual, to parents, and to the community to which the individual belongs" (p.224). Once a target is in place, feedback is needed if the level of necessary application is to be maintained. This requires teachers to put their judgement on the line, something which they are often unwilling to do. Much of what has been written about target setting, according to Flecknoe, is therefore "in the nature of unsupported assertion" (p.226).

Fielding (1999) is even more forthright in his criticisms and concern about the concept of target-setting, because of the "doubtful quality of the evidence base" (p.277). It has become the "viagra of economic and educational under-performance" (p.277), not least because it satisfies "the prurience of an increasingly promiscuous accountability" (p.277). This raises yet another largely unexamined assumption about target setting. How valid is it to use such procedures for accountability purposes? What are the effects on the validity and reliability of target-setting procedures if they are being used simultaneously both for raising standards and for making judgements? Evidence from the rest of the public sector, not least in the National Health Service, suggests that as the accountability stakes are raised, the tendency for targets to be manipulated, and for data upon which they are based to be manipulated, increases. Fielding proceeds to ask a host of pertinent questions. Can efficiency and legitimacy be regarded as one and the same thing? Can target-setting become an end in itself? Do we merely

end up "measuring the measurable" (p.280) and misrepresenting "the immeasurable" (p.280)? What are the aspects of education which we hold dear, but are actually quite elusive? What are the effects on creativity of this kind of utilitarian rationality? What are the real effects on performance?

Fielding takes the view that targets, and particularly the fear of not meeting them, causes teachers to retreat into technical proficiency and students into habit and mimicry. The idea that teachers and students "own" targets may be a useful "mantra" but in practice schools and students will "feel the weight and wisdom of a reality defined largely by others, albeit for what is judged to be their own and, hopefully, the greater good" (p.281). At a broader level, the apparent transparency of target-setting, which makes it so seductive within this kind of performative thinking, "runs a substantial risk of becoming prey to reductionism and distortion" (p.282). Used for accountability purposes, target-setting outcomes become particularly distorted and a whole new set of values and assumptions come into play. Above all, says Fielding, the gulf between "the minutiae of performative paranoia" (p.286) and the "complex reality of a lived partnership" (p.286) within schools is stark, and "clues for meaning are trodden underfoot in the scramble for performance" (p.286). Target-setting is not just about oft-cited teaching to the test. The very nature of the relationship between teacher and taught is distorted, and subtlety, passion and integrity are devalued by the "metallic language of the market" (p.287).

Whilst the DFEE (1996) report embodies the philosophy of a Conservative government, we not only see no change in this philosophy three years later, with a Labour government in power, but rather an intensification of the concept of educational targets and economic competitiveness. The DFEE (1999) is forthright from the outset, in a foreword by the Secretary of State for Education, David Blunkett: "lack of skills and low educational achievement have been weaknesses of the British economy for more than a generation" (p.1) and we can only overcome this by "the development of

education and skill levels sufficient to achieve economic success and an inclusive society" (p.1). There is, however, more than an element of tension, if not contradiction here. Are these "twin aims" of "greater competitiveness and improving quality of opportunity" (p.1) not at odds with each other? Competition produces winners and losers. That is the fundamental basis upon which competition is built. Quite how a society that encourages an education system which is inextricably linked, through target-setting, to global competition, can also be assumed to be pursuing the goals of equality of opportunity, and an inclusive society, remains unresolved. In the 1990s, according to the report, the performance of pupils has improved, but in recent years "the rate of increase has slowed considerably" (p.3). This raises an interesting question: even if targets can be shown to be effective in raising standards, just *how far* can standards be raised? Is there a finite limit, a plateau beyond which we cannot go? Or do we accept a little examined, taken-for-granted-assumption, that there is no ceiling to what any individual, or any school can achieve, given the right leadership, conditions, teachers, effort, vision, or any of the other factors cited in research on standards? Just as important are questions of cost. Targets may well focus our attention more specifically on certain actions and types of behaviour, but it is tempting to wonder what loses out in the process. We may single-mindedly focus our efforts on academic achievement in schools, but what sort of things are likely to be neglected as a result? We do, after all, only have a finite amount of energy and attentional capacity. If we concentrate this on a more narrow focus, it rather suggests that other things which we once paid attention to will no longer concern us. This assumption, like so many in government publications about target-setting, remains unexamined.

National Targets are set in this document. Whilst 65% of 11 year olds reached at least Level 4 in Key Stage 2 English in 1998, this must rise to 80% by 2002. In Mathematics the figures are 59% and 75% (despite the fact that in 1998, the figure for Mathematics was actually 3% lower than it was in 1997). At GCSE, in 1998 46.3% of students achieved five or more A*-C

grades and by 2002 this figure must be 50%. No targets are set for Key Stage 3 in this document and it is tempting to ask why.

Further concerns about target-setting are also raised by Tikly et al (2000). The assumption that institutional and individual needs "are necessarily concurrent" (p.12) clearly needs to be opened up for debate and one of the key questions is therefore "whether the setting of whole school targets is supportive of the learning of each child" (p.12). Targets also have important effects on our wider conceptions of education and the purposes we assume that it has in society. "The reification of targets as a central goal of education, tends to obscure and even stifle debate on whether the targets themselves are consistent with acceptable educational objectives" (p.13). Concerns were expressed by teachers who were interviewed as part of this study with regard to making predictions, for which teachers will be held responsible and Tikly et al feel that there is a significant risk that this will lead to the same kind of defensive professional practice as we see emerging in the medical profession. In terms of setting targets, for instance, this presumably means that any institution will set targets which are well within its capacity to achieve, because to do otherwise would potentially be setting the institution up for a very public accusation of failure. Individual teachers, presumably will operate in much the same kind of way. If this is what happens, in practice, we must ask ourselves what targets are actually achieving. Despite all the talk about targets being "challenging" it is entirely possible that targets are anything but challenging. You don't set yourself up for failure if you care about survival. At a more cynical level, it is clearly tempting to wonder, given what we have seen in other areas of public life, to what extent easily achievable targets are put in place with the specific intention of demonstrating to the world how effective the performance of an individual, or institution, actually is. Fielding (1998) observes that target-setting "tends to encourage manipulation of figures rather than genuine improvement" (p.42) and that it can even "create illusions of progress" (p.42). The "most likely outcome of widespread target-setting is widespread

statistical manipulation" (p.42). In a nutshell, the fundamental objective of targets, namely to raise standards, can easily be subverted, and even replaced by the rather more tawdry, if not sinister, objective of influencing or "managing" public perceptions. What all this tends to produce is a kind of bureaucracy in which there is a real danger of "resource wastage, unwillingness to take risks and possibly disincentives for students" (p.13). The view of Tikly et al is that moving towards the idea of defining targets for individual students can be useful in alleviating some of these dangers but it is difficult to see how this can be the case. Student psychology is no different from the psychology of anybody else. The same dangers exist for students with individual targets as we have seen in the wider field of education. Furthermore, as we will see in this study, there is evidence that when they fail to achieve a target, the blame tends to be levelled at the school, or the teacher, rather than at themselves.

WIDER ASSESSMENT ISSUES

Given the pervasive nature of assessment of every conceivable kind in our education system, let alone in society at large, it is curious how little attention has been paid, until recently, to the *impact* of assessment on the recipients of such judgements. As we saw in the discussion of validity by Messick, and Crooks et al, the consequences, or the impact, of any form of educational measurement has significant implications for the validity of that process. Broadfoot (2000) sees assessment worldwide as having become "the unquestioned arbiter of value" (preface) and that this dominance has gone almost entirely unchallenged, despite the very obvious social cost of assessment in terms of disillusionment and disenchantment. Broadfoot feels that this tends to inhibit the development of creativity in the individual because "the rewards go to the dogged and the dutiful, the convergent and the conforming" (preface). Assessment systems are increasingly becoming "mechanisms of power and control" (preface) and therefore "a new means of legitimating social inequalities" (preface). Broadfoot sees assessment,

therefore, as "useful" (though she does not specify how it is useful), but she also sees the potential "contagion of 'performativity' and its symptoms of pervasive judgement and comparison" (preface) as carrying significant dangers which we need to recognise and take account of. Filer (2000a) also sees real social dangers in the way that assessment is used, suggesting that it "underpins and legitimizes the reproduction of social and power elites in modern society" (p.8). Given that developed societies introduced procedures such as examinations and testing in an attempt to move away from the concept of privilege and power being dependent on the position of the individual in the social hierarchy, and produce a more meritocratic way of organising society, this presents us with something of a dilemma if Filer is right. If a meritocratic system of assessment reproduces social inequalities how do we proceed? If we don't reward people on the basis of merit, how do we reward people? Filer might argue, perhaps, that all we can do is to try to mitigate the worst effects of the social impact of assessment procedures through greater awareness of the issues involved. It is reasonable to presume that the majority of professionals involved in education would support this idea. This still leaves the question of the impact on the *validity* of such procedures.

Broadfoot and Pollard (2000), in charting changes in primary education in England in the last decade, see it as reflecting the changes in "power relations and modes of control of modern societies" (p.12). This has caused teachers to become much more instrumental in their approach to their jobs in the classroom and has caused students to become much more performance orientated. "The achievement of targets is "becoming a new vehicle for satisfaction and professional fulfilment" (p.22) and as this utilitarian attitude towards education becomes more embedded in the system, "it may become increasingly difficult to encourage many pupils to adopt positive attitudes towards learning" (page 22). This seems to be a particular risk for those students who are unable to achieve their targets for whatever reason. At a broader level therefore education has become

"subjugated as never before to the educational assumptions of an earlier, modernist, industrial age" (p.25). This produces a kind of social technology of testing (Filer 2000b) which embodies all sorts of hidden social and cultural values, assumptions and expectations.

Madaus and Horn (2000) raise an interesting question about this concept of testing as a form of technology in society. In all other forms of technology we have substantial procedures in place to protect the public from danger, abuse, or exploitation. We spend extremely large sums of money testing everything from aircraft to clinical drugs, not just before we unleash these technologies on the public, but also throughout the active shelf-life of the product. Yet assessment procedures, which have equally as fundamental consequences for individuals, are the subject of relatively little attention or concern beyond the educational community itself. What we see, in effect, is the educational community policing itself and policy makers able to implement developments with little in the way of public protection. An individual undergoing a new medical procedure has reasonable grounds for assuming that this procedure has been thoroughly tested and evaluated and, as we see in the medical world, patients increasingly expect any potential risks to be quantified and identified to them. Given the impact of assessment procedures on individuals in society, it is surely a matter of concern, as Madaus and Horn suggest, that so little exists in the way of public protection especially when "many people are simply not aware of the control these hidden systems can exert on their lives" (p.58). We therefore need "some form of more active and independent monitoring" (p.60). Whilst these influences may remain hidden at the moment, it is clear from the comments of Linn (1993), as we have seen, with regard to the use of the legal system in the USA in recent years, that these influences and consequences are likely to become more widely known, and debated, in the future, and this poses a significant problem for the public perception, and therefore the validity, of all assessment procedures, not just those used in education.

A more direct consequence of assessment procedures is examined by Hanson (2000), who argues that "tests often produce the characteristics they purport to measure" (p.68). Tests do not measure independently existing realities, they create them, and thereby become an example of the kind of simulacrum described by Baudrillard because they "take on a reality more palpable than the things they represent" (p.70). The signifier becomes more real than what it signifies. Because tests "construct human traits rather than measuring them" (p.71) individuals come to think of themselves in terms of what the tests tell them. (This is clearly reminiscent of Marx's (1971) concept that our social consciousness is not something we hold independently of society, but is actually constituted *by* our social experience). Intelligence testing, for Hanson, is a classic example of how tests enable "a new mental category to swim into the social consciousness" (p.77) and we then classify ourselves in terms of these categories. This sets up all sorts of expectations and assumptions in the test taker as well as in those using the outcomes of the test. As a result of these effects, Hanson takes the view that Foucault is correct in maintaining that what examinations of all kinds give rise to is an "increasingly total network of surveillance and control" (p.79) which enables society to control the very way in which individuals think about themselves.

These dangers even exist in formative assessment according to Pryor and Torrance (2000) because the learner's knowledge and classroom identity is inextricably bound up with social interactions in the classroom. Students make all sorts of judgements about the validity and importance of various types of knowledge and viewpoints as a result of these social interactions. Their perceptions of themselves are then changed by the way they see their contributions being received by the teacher, and by other pupils. Teachers imply comparisons, even if they strive not to make them explicit. Because the criteria for making judgements in formative assessment remain hidden (even if it is possible to identify what they are) students inevitably draw

their own inferences about why their work is judged in the way that it is. As a result of all these contextual effects, the conclusion of many writing in this area is that the validity of assessment procedures is in very serious doubt. Berlak (2000) sees an "overwhelming body of evidence" (p.193) which suggests that "standardized measures of educational performance are a mirage" (p.193). Torrance (2000) sees the idea of trying to "provide universally correct answers" (p.185) as inadequate to the educational task. Any postmodernist conception of assessment, for Torrance, "would attempt to respond to the fragmented, individualistic, multicultural and post-colonial culture which we now inhabit" (p.186). Some might view this as a relatively tall order. Cooper and Dunne (2000) argue that given the nature of the society in which we find ourselves we simply cannot assume that in any assessment task there is only one single problem embedded in the text, "but rather a whole range of possible problems which might be constructed by children" (p.102). In an investigation of tests in Mathematics they found that even at the age of fourteen a considerable minority of children have problems "reading the intentions of the test designers" (p.107) in a way which enables them to demonstrate the skills which the test purports to measure. This is more than just lack of reading skill. It is about the ability to read intentions, which are often culturally bounded in ways which we often fail to appreciate. If Cooper and Dunne are correct, the validity of many assessment procedures for these types of children is seriously threatened.

Wiggins and Tymms (2000) see threats to school level data as a result of making performance data publicly available. In a comparison of test results for primary schools which are published in England, but not in Scotland, they conclude that such publication "appears to have significant dysfunctional effects on the long term management and organisation of schools" (p.2) leading to a narrowing effect on the curriculum, a concentration of resources on borderline children, and an increasing blame culture. The drive to measure performance quickly and efficiently leads to the use of "proxy indicators of overall performance" (p.5) and thus to a

whole series of unintended consequences such as moving the goal-posts, changing the rules and re-classifying categories of data. Once again, we see the validity of the data we are likely to be working with to be in some doubt. Add to this the consequential effects of decisions and judgements made on the basis of such data, and all the ingredients are in place for the integrity of the entire system of assessment to be called into question. Whilst confidence in such data may only currently be in crisis *within* the profession of education itself, it is tempting to wonder what the consequences might be if this crisis spreads to the public domain, and what the consequential effects will be for students, and for the very nature of assessment itself.

As we have seen, assessment is now inextricably linked with the concept of raising standards, particularly in the context of global economic competition, and this again has an important impact on the validity of assessment procedures. Brown (1998) raises serious questions about the validity of many of the so-called international comparisons of educational effectiveness, citing fundamental sampling problems, age differentials due to some students being held back for a year in some countries, schools refusing to take part in various countries, and the fact that in many developing countries only a proportion of students attend school anyway. For Slee and Weiner (1998) this kind of effectiveness approach leads to a "manufactured crisis" (p.6) using statistical techniques which are no more than a "ready reckoner" (p.6) of benchmarks against which to measure performance. Such an approach "bleaches context from its analytic framework" (p.5), producing a "set of platitudes" (p.5) and a "litany of alleged failures in state schooling" (p.1). For Hamilton (1998) this produces a "cash and carry market for educational panaceas" (p.16) based on " a cluster of assumptions, claims and propositions" (p.14) and which gives rise to a hankering after "tablets, magic bullets, and smart missiles" (p.14). School effectiveness for Hamilton "stands at the intersection between educational research and social engineering" (p.13) and has therefore "forced the debate and over-ruled the understanding" (p.19). Teachers are

thus gradually disempowered according to Rea and Weiner (1998), despite the claims of the effectiveness movement to the contrary, because the approach embodies models of change which become "bureaucratic, and managerial, administered through development plans, programmes, targets, outputs and measures of success and failure" (p.27). School effectiveness therefore becomes a tool of "control" not of "liberation" (p.28). What this all produces is a mechanistic performativity in which our social experience is explained in terms of "order, structure, function, cause and effect" (p.73) thereby excluding the "mobile, complex, ad hoc, messy and fleeting qualities of our lived experience" (p.73) according to Ball (1998). The human sciences have thus been accorded the kind of neutral status usually accorded to the natural sciences, and this, for Ball, makes them potentially dangerous and debilitating within a political or educational context, not least because it leads to the "commodification of education" (p.74) and a "technology of control" (p.74). What we really need, according to Lingard et al (1998) is to recognise that schools are not industrial machines, but complex ecosystems in which we need to make allowances for "hybridity, idiosyncrasy, difference, and indeed local and pragmatic accidents" (p.94). As Riddell et al (1998) put it, "recognition of the complexity of the classroom context may be inconvenient for those looking for straightforward comparisons among schools" (p.172) but unless we recognise the considerable differences between individual schools and classrooms, we are unlikely to be very effective in our attempts to bring about improvement.

ACCOUNTABILITY

Similar problems can be identified with regard to the issue of accountability. There was a time when assessment was conceived of as a tool to measure the performance of students. In the 1990s it became a tool to *measure* the performance of schools, and then a tool to *compare* schools. More recently, as we have seen, it has become a tool for making international comparisons,

and for the professional appraisal of teachers. The consequential effects of all this in terms of the validity chain of Crooks et al (1996) are considerable, not least because accountability procedures can so easily become procedures of control, further distorting perceptions of, and attitudes towards, the processes of assessment. What we see is an increasing weakening of the professional and partnership models of accountability described by Halstead (1994) and an increasing emphasis on the consumerist, chain of responsibility, and central models, in other words an increasing tendency to erode the professional autonomy and initiative of teachers. Scott (1994a) argues that accountability and assessment have become ever more inextricably linked since the 1988 Education Reform Act, and that whilst the concept of accountability is inevitably always going to be a contested concept, it is nevertheless incumbent upon us to try to reach some kind of adequate definition, otherwise teachers just find themselves caught in an ever-changing arena of conflicting values and assumptions. The critical issue for Scott (1994b) is that "data collected about schools and education systems can be useful only if they accurately represent how schools and systems actually function" (p.167). As we have seen, the accuracy of such representations is not only open to question, but it has important consequential effects on the behaviours that give rise to the data in the first place.

Yet despite these fundamental problems, the balance between teacher autonomy and public accountability continues to be eroded in favour of greater accountability and as McCulloch (2001) has suggested many observers agree with the perception of teachers themselves that they have "been systematically deskilled and deprofessionalised" (p.108). The introduction of literacy and numeracy hours has only tended to intensify this feeling whilst the far-reaching reforms in teacher training in the last twenty years has, according to Furlong (2001) largely abandoned any concept of the teacher as expert, or as professional practitioner, in favour of "common

standards and procedures" (p.131) where the "voice of the individual teacher is almost wholly silenced" (p.132). What this all tends to lead to is a system of accountability founded more or less exclusively on what is directly measurable or observable. Given the problems which we have identified in terms of what value-added data actually does measure, the concept of using such data for accountability purposes becomes extremely problematic. Using target-setting procedures constructed on the basis of value-added assumptions only compounds these problems further.

SUMMARY

Despite the considerable problems which surround the concept of assessment, let alone procedures such as value-added, and target-setting, as we have seen, successive legislative acts by government have forced schools to use such procedures, and to make judgements about students, teachers, and their own institutions, on the basis of such procedures. Given that assertions made by government are based on very large national data sets, which produce many of the types of distorting effects, as a result of aggregation, that we have seen, an obvious question to ask is what the data reveal over a five-year time-period within one institution, and how these data sets might compare with national data. In particular we need to know how valid and reliable different types of baseline data might be, and how valid and reliable linear regression procedures, recommended by government, might be, both in calculating value-added, and in setting targets. It is in the next chapter that we examine this issue in some detail.

CHAPTER 4

HIDDEN PATTERNS AND TRENDS

ANALYSIS AND DISCUSSION OF THE NUMERICAL DATA

As we saw in Chapter 3, SCAA (1997a), in their report on the *Value-added National Project,* defined value-added as the difference between a statistically predicted performance (based on prior attainment) and actual performance. We noted how both the validity and reliability of these procedures depend crucially on the validity and reliability of the baseline data which is utilised because value-added procedures involve trying to identify patterns over a period of time (usually at least two years) in order to compare the performance of institutions, or even classes, or individual teachers.

One way of measuring the validity of the baseline data is to look at both the strength, and the nature, of the relationship between such data and subsequent performance data. Two statistical values are useful here. The first is the R-value which simply tells us what the strength of the relationship is between the baseline data and the subsequent performance data. It is important to be clear that this tells us nothing about *causation.* If we take the R-value and square this number (R^2) we can derive what Ferguson and Takane (1989) call the coefficient of determination. This value reveals the extent to which the variation in the data set for Key Stage 3 is accounted for by the variation in either the CAT data, or the Key Stage 2 data. This value is expressed on a scale from 0 to 1. An R value of 0.7, though recommended as a good basis for value-added assumptions by SCAA (1997a), when squared, gives us a value of 0.49. In other words only 49% of the variation in the second data set can be accounted for by the variation in the first data set. This means that 51% of the variation is due to some other factor, or factors, and furthermore these factors, or any

combination of such factors, could vary considerably from one institution to another. We also noted that SCAA (1997a) was sufficiently confident about Key Stage 2 test outcomes that it was able to recommend that from 1998 it would be possible to predict Key Stage 3 outcomes from Key Stage 2 outcomes. Measuring the reliability of baseline data involves looking at the pattern of relationships with later measures of performance over a period of time, in order to ascertain whether there is some kind of regular pattern in the relationships, or whether these relationships fluctuate in ways which are difficult to explain, or predict.

Since, as we have seen in Chapter 3, the NFER CAT data has a demonstrably high level of reliability over time, and since this also represents a standardised test which does not change year on year, this provides us with a useful baseline from which to make judgements about Key Stage 3 performance over time, and to make comparisons with the relationships between Key Stage 2 and Key Stage 3 data. (The raw data for these comparisons can be found in Appendices 7-26).

VALUE-ADDED DATA: ANALYSIS AND DISCUSSION

The population sizes for the analyses which follow were:

1994 N = 131 1995 N = 153 1996 N = 142

1997 N = 166 1998 N = 143

If we look first at the relationships for English, two features of this data set are fairly obvious, as we can see in Table 4.1. The first is that there is a wide variation in these coefficients from 0.40 to 0.87 across the various sub-sets of data. Secondly there is considerable fluctuation across the five year period for each sub-set of data. Overall, the relationship between CAT data and Key Stage 3 data is generally stronger than that between Key Stage 2 data and Key Stage 3 data. Many of the coefficients are below the 0.7 value which schools have been advised to take as a valid basis upon which to make value-added calculations and assumptions.

TABLE 4.1

REGRESSION (R) OF CAT AND KEY STAGE 2 DATA WITH KEY STAGE 3 DATA FOR ENGLISH, INTAKE YEARS 1994-1998

Subject	Regression	1994	1995	1996	1997	1998
English	CATVerbal/KS3 TA level	0.87	0.71	0.79	0.79	0.75
	CAT Verbal/KS3 Test level	0.77	0.69	0.75	0.73	0.68
	CAT Quantitative/KS3 TA level	0.72	0.66	0.70	0.67	0.57
	CAT Quantitative/ KS3 Test level	0.56	0.49	0.67	0.60	0.55
	CAT Non-verbal/KS3 TA level	0.69	0.57	0.58	0.68	0.62
	CAT Non-verbal/KS3 Test level	0.51	0.40	0.51	0.55	0.48
	CAT Average/ KS3 TA level	0.85	0.72	0.78	0.71	0.73
	CAT Average/ KS3 Test level	0.69	0.61	0.73	0.63	0.65
	KS2 TA/ KS3 TA level	N/A	0.64	0.69	0.73	N/A
	KS2 TA/KS3 Test level	N/A	0.56	0.68	0.66	N/A
	KS2 Test level/KS3 TA level	N/A	0.58	0.69	0.64	0.61
	KS2 Test level/KS3 Test level	N/A	0.49	0.71	0.62	0.56
	KS2 Average test level/KS3 TA level	N/A	0.64	0.75	0.71	0.67
	KS2 Average test level/KS3 Test level	N/A	0.51	0.73	0.69	0.62

If we take these values and convert them into R^2 values, we can get a stronger sense of the nature of these relationships, as can be seen in Table 4.2. Looked at in this way, it is clear that CAT data, and Key Stage 2 data fail to account for the variation in the Key Stage 3 data, to a significant extent, on many of the data subsets. Typically, some 50%-70% of the variation in Key Stage 3 data is not accounted for by either CAT data or Key Stage 2 data.

TABLE 4.2

REGRESSION (R^2) OF CAT AND KEY STAGE 2 DATA WITH KEY STAGE 3 DATA FOR ENGLISH, INTAKE YEARS 1994-1998

Subject	Regression	1994	1995	1996	1997	1998
English	CATVerbal/KS3 TA level	0.76	0.50	0.62	0.62	0.57
	CAT Verbal/KS3 Test level	0.60	0.47	0.56	0.53	0.46
	CAT Quantitative/KS3 TA level	0.52	0.43	0.49	0.45	0.32
	CAT Quantitative/ KS3 Test level	0.31	0.24	0.45	0.36	0.30
	CAT Non-verbal/KS3 TA level	0.48	0.33	0.34	0.46	0.39
	CAT Non-verbal/KS3 Test level	0.26	0.16	0.26	0.30	0.23
	CAT Average/ KS3 TA level	0.73	0.52	0.61	0.50	0.54
	CAT Average/ KS3 Test level	0.47	0.37	0.53	0.40	0.42
	KS2 TA/ KS3 TA level	N/A	0.41	0.47	0.54	N/A
	KS2 TA/KS3 Test level	N/A	0.31	0.46	0.44	N/A
	KS2 Test level/KS3 TA level	N/A	0.34	0.47	0.41	0.37
	KS2 Test level/KS3 Test level	N/A	0.24	0.50	0.38	0.31
	KS2 Average test level/KS3 TA level	N/A	0.41	0.57	0.51	0.45
	KS2 Average test level/KS3 Test level	N/A	0.26	0.54	0.48	0.38

If we conduct the same kind of analysis of data sets for Mathematics, using the R-value, we find a rather different picture, as can be seen in Table 4.3. These coefficients are clearly much better than those for English, ranging from 0.73 to 0.90. The figures are much higher for the CAT data than for the Key Stage 2 data, with the CAT average score producing higher values than the CAT Quantitative score. In general the strength of the relationship of both CAT data and Key Stage 2 data, with Key Stage 3 data, seems to be weakening a little over the five years of the study, though the type of fluctuations found in the English data are not apparent here.

TABLE 4.3

REGRESSION (R) OF CAT AND KEY STAGE 2 DATA WITH KEY STAGE 3 DATA FOR MATHEMATICS, INTAKE YEARS 1994-1998

Subject	Regression	1994	1995	1996	1997	1998
Maths	CAT Verbal/KS3 TA level	0.77	0.75	0.77	0.77	0.76
	CAT Verbal/KS3 Test level	0.75	0.73	0.80	0.78	0.75
	CAT Quantitative/KS3 TA level	0.85	0.81	0.84	0.85	0.75
	CAT Quantitative/ KS3 Test level	0.85	0.81	0.82	0.87	0.75
	CAT Non-verbal/KS3 TA level	0.80	0.79	0.71	0.79	0.77
	CAT Non-verbal/KS3 Test level	0.81	0.78	0.75	0.79	0.77
	CAT Average/ KS3 TA level	0.90	0.88	0.87	0.81	0.86
	CAT Average/ KS3 Test level	0.89	0.87	0.89	0.82	0.85
	KS2 TA/ KS3 TA level	N/A	0.71	0.74	0.81	N/A
	KS2 TA/KS3 Test level	N/A	0.71	0.76	0.82	N/A
	KS2 Test level/KS3 TA level	N/A	0.78	0.73	0.77	0.71
	KS2 Test level/KS3 Test level	N/A	0.81	0.73	0.77	0.67
	KS2 Average test level/KS3 TA level	N/A	0.77	0.78	0.77	0.77
	KS2 Average test level/KS3 Test level	N/A	0.77	0.80	0.78	0.75

The R^2 values for Mathematics, as can be seen in Table 4.4 indicate that a great deal more of the variation in the Key Stage 3 data can be explained in terms of the variation in either CAT data or Key Stage 2 data than was the case for the English data. Overall, somewhere between 60% and 70% of the Key Stage 3 variation can be explained in this way, though for some data subsets this figure is either slightly lower or slightly higher.

TABLE 4.4

REGRESSION (R^2) OF CAT AND KEY STAGE 2 DATA WITH KEY STAGE 3 DATA FOR MATHEMATICS, INTAKE YEARS 1994-1998

Subject	Regression	1994	1995	1996	1997	1998
Maths	CAT Verbal/KS3 TA level	0.60	0.57	0.60	0.60	0.58
	CAT Verbal/KS3 Test level	0.57	0.54	0.64	0.61	0.57
	CAT Quantitative/KS3 TA level	0.72	0.66	0.70	0.72	0.57
	CAT Quantitative/ KS3 Test level	0.72	0.66	0.67	0.75	0.57
	CAT Non-verbal/KS3 TA level	0.64	0.63	0.50	0.62	0.59
	CAT Non-verbal/KS3 Test level	0.65	0.61	0.57	0.63	0.59
	CAT Average/ KS3 TA level	0.81	0.77	0.75	0.65	0.74
	CAT Average/ KS3 Test level	0.80	0.76	0.79	0.67	0.73
	KS2 TA/ KS3 TA level	N/A	0.50	0.55	0.65	N/A
	KS2 TA/KS3 Test level	N/A	0.50	0.58	0.67	N/A
	KS2 Test level/KS3 TA level	N/A	0.61	0.54	0.59	0.51
	KS2 Test level/KS3 Test level	N/A	0.65	0.54	0.60	0.45
	KS2 Average test level/KS3 TA level	N/A	0.60	0.61	0.59	0.60
	KS2 Average test level/KS3 Test level	N/A	0.60	0.64	0.61	0.57

For Science, a rather more complex pattern of relationships becomes apparent using the R-value, as can be seen in Table 4.5. The coefficients for CAT data and Key Stage 3 data range from 0.61 to 0.82 though there is quite a lot of fluctuation over the five year period. Key Stage 2 data, with the exception of the Key Stage 2 average test level, seems to produce poor figures, ranging from 0.46 to 0.74. For some reason which is not entirely clear, the Key Stage 2 average scores produce higher coefficients, ranging from 0.69 to 0.81, though these are a lot weaker than the figures for the CAT average which ranges from 0.73 to 0.82.

TABLE 4.5

REGRESSION (R) OF CAT AND KEY STAGE 2 DATA WITH KEY STAGE 3 DATA FOR SCIENCE, INTAKE YEARS 1994-1998

Subject	Regression	1994	1995	1996	1997	1998
Science	CAT Verbal/KS3 TA level	0.78	0.81	0.75	0.79	0.73
	CAT Verbal/KS3 Test level	0.79	0.78	0.79	0.84	0.77
	CAT Quantitative/KS3 TA level	0.70	0.68	0.66	0.70	0.63
	CAT Quantitative/ KS3 Test level	0.69	0.71	0.69	0.71	0.61
	CAT Non-verbal/KS3 TA level	0.69	0.71	0.62	0.67	0.66
	CAT Non-verbal/KS3 Test level	0.64	0.70	0.66	0.70	0.65
	CAT Average/ KS3 TA level	0.81	0.82	0.75	0.73	0.77
	CAT Average/ KS3 Test level	0.79	0.82	0.79	0.75	0.77
	KS2 TA/ KS3 TA level	N/A	0.46	0.63	0.62	N/A
	KS2 TA/KS3 Test level	N/A	0.54	0.66	0.74	N/A
	KS2 Test level/KS3 TA level	N/A	0.56	0.71	0.62	0.58
	KS2 Test level/KS3 Test level	N/A	0.56	0.72	0.74	0.62
	KS2 Average test level/KS3 TA level	N/A	0.74	0.76	0.74	0.69
	KS2 Average test level/KS3 Test level	N/A	0.75	0.77	0.81	0.74

The R^2 values for Science, which can be seen in Table 4.6 show that whilst the CAT data can account for around 50% to 60% of the variation in Key Stage 3 data, the Key Stage 2 data can only account for about 30% to 40%, with the Key Stage 2 average test level accounting for about the same level of variation as the CAT data. There are individual fluctuations for various sub-sets of the data, some of which are extremely low.

TABLE 4.6

REGRESSION (R^2) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 DATA IN SCIENCE, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
Science	CAT Verbal/KS3 TA level	0.61	0.65	0.56	0.63	0.54
	CAT Verbal/KS3 Test level	0.63	0.61	0.62	0.71	0.60
	CAT Quantitative/KS3 TA level	0.49	0.47	0.43	0.49	0.40
	CAT Quantitative/ KS3 Test level	0.48	0.50	0.47	0.51	0.37
	CAT Non-verbal/KS3 TA level	0.47	0.50	0.39	0.45	0.44
	CAT Non-verbal/KS3 Test level	0.41	0.49	0.43	0.49	0.42
	CAT Average/ KS3 TA level	0.66	0.67	0.57	0.53	0.59
	CAT Average/ KS3 Test level	0.63	0.67	0.62	0.57	0.59
	KS2 TA/ KS3 TA level	N/A	0.21	0.40	0.39	N/A
	KS2 TA/KS3 Test level	N/A	0.29	0.44	0.55	N/A
	KS2 Test level/KS3 TA level	N/A	0.31	0.51	0.39	0.34
	KS2 Test level/KS3 Test level	N/A	0.31	0.52	0.55	0.38
	KS2 Average test level/KS3 TA level	N/A	0.55	0.58	0.55	0.48
	KS2 Average test level/KS3 Test level	N/A	0.57	0.60	0.65	0.55

English, Mathematics and Science are the only three subjects in the National Curriculum which are tested. For all the other subjects, teacher assessments are used. If we look at the R-values for Design Technology (Table 4.7) the most remarkable feature of this analysis is how low the coefficients are, ranging from 0.22 to 0.64. The figures for the CAT data are substantially better than those for the Key Stage 2 data, ranging from 0.44 to 0.64 for the CAT data, and from 0.22 to 0.61 for the Key Stage 2 data. Despite these differences there is some improvement across all the data sub-sets for the five years, though there are also some anomalous fluctuations too.

TABLE 4.7

REGRESSION (R) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN DESIGN TECHNOLOGY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
DT	CAT Verbal/KS3 TA	0.49	0.44	0.55	0.56	0.57
	CAT Quantitative/KS3 TA	0.46	0.45	0.57	0.57	0.52
	CAT Non-verbal/KS3 TA	0.52	0.41	0.48	0.54	0.60
	CAT Average/KS3 TA	0.56	0.48	0.59	0.57	0.64
	KS2 English test level/KS3 TA	N/A	0.22	0.50	0.41	0.50
	KS2 Maths test level/KS3 TA	N/A	0.28	0.56	0.47	0.55
	KS2 Science test level/KS3 TA	N/A	0.24	0.53	0.44	0.47
	KS2 Average test level/KS3 TA	N/A	0.30	0.61	0.52	0.60

Given how low the R-values are for Design Technology, it is inevitable that the R^2 values (Table 4.8) will be very low. Only some 20% to 30% of the variation at Key Stage 3 can be accounted for by either CAT data, or Key Stage 2 data, and some of these values are exceedingly low, suggesting that some 90% or more of the variation cannot be explained by the baseline data.

TABLE 4.8

REGRESSION (R^2) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN DESIGN TECHNOLOGY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
DT	CAT Verbal/KS3 TA	0.24	0.19	0.30	0.32	0.33
	CAT Quantitative/KS3 TA	0.21	0.20	0.32	0.32	0.27
	CAT Non-verbal/KS3 TA	0.27	0.17	0.23	0.29	0.36
	CAT Average/KS3 TA	0.31	0.23	0.35	0.33	0.41
	KS2 English test level/KS3 TA	N/A	0.05	0.25	0.17	0.25
	KS2 Maths test level/KS3 TA	N/A	0.08	0.31	0.22	0.30
	KS2 Science test level/KS3 TA	N/A	0.06	0.28	0.19	0.22
	KS2 Average test level/KS3 TA	N/A	0.09	0.37	0.27	0.36

As with Design Technology, Information and Communication Technology shows poor R-values generally for both CAT data and Key Stage 2 data, with the odd exception (Table 4.9). Perhaps the most notable feature of this analysis is the considerable fluctuations, year on year, in the coefficients,

which can show differences of between 0.20 and 0.30 in successive years, for both CAT and Key Stage 2 data.

TABLE 4.9

REGRESSION (R) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN INFORMATION AND COMMUNICATION TECHNOLOGY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
ICT	CAT Verbal/KS3 TA	0.49	0.56	0.82	0.42	0.63
	CAT Quantitative/KS3 TA	0.45	0.69	0.74	0.46	0.58
	CAT Non-verbal/KS3 TA	0.40	0.66	0.61	0.45	0.57
	CAT Average/KS3 TA	0.50	0.71	0.81	0.46	0.67
	KS2 English test level/KS3 TA	N/A	0.41	0.66	0.36	0.51
	KS2 Maths test level/KS3 TA	N/A	0.55	0.64	0.39	0.52
	KS2 Science test level/KS3 TA	N/A	0.32	0.66	0.33	0.55
	KS2 Average test level/KS3 TA	N/A	0.52	0.74	0.42	0.62

As with Design Technology, the R^2 values are inevitably very low for Information and Communications Technology (Table 4.10). The large variations in the data over time are, of course, repeated here, and typically only about 30% to 40% of the data variation can be explained by either CAT data, or Key Stage 2 data. Once again, a few of the coefficients suggest that some 90% of the variation in the data cannot be accounted for by the baseline data.

TABLE 4.10

REGRESSION (R^2) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN INFORMATION AND COMMUNICATION TECHNOLOGY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
ICT	CAT Verbal/KS3 TA	0.24	0.31	0.67	0.18	0.40
	CAT Quantitative/KS3 TA	0.20	0.47	0.55	0.21	0.34
	CAT Non-verbal/KS3 TA	0.16	0.44	0.37	0.20	0.33
	CAT Average/KS3 TA	0.25	0.50	0.66	0.21	0.45
	KS2 English test level/KS3 TA	N/A	0.17	0.43	0.13	0.26
	KS2 Maths test level/KS3 TA	N/A	0.30	0.41	0.15	0.27
	KS2 Science test level/KS3 TA	N/A	0.10	0.43	0.11	0.30
	KS2 Average test level/KS3 TA	N/A	0.27	0.55	0.18	0.39

The R-values for History (Table 4.11) are clearly somewhat better than those for either Design Technology, or Information and Communication Technology. The coefficients for CAT data range from 0.45 to 0.80 whilst those for Key Stage 2 data range from 0.35 to 0.75. However, as we have seen in earlier data sets, there are some large fluctuations in several places from one year to the next with single sub-sets of the data changing by as much as 0.20 from one year to the next.

TABLE 4.11

REGRESSION (R) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN HISTORY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
History	CAT Verbal/KS3 TA	0.77	0.57	0.79	0.80	0.70
	CAT Quantitative/KS3 TA	0.55	0.56	0.71	0.72	0.57
	CAT Non-verbal/KS3 TA	0.56	0.45	0.64	0.70	0.59
	CAT Average/KS3 TA	0.70	0.59	0.80	0.76	0.70
	KS2 English test level/KS3 TA	N/A	0.35	0.66	0.66	0.57
	KS2 Maths test level/KS3 TA	N/A	0.40	0 63	0.61	0.57
	KS2 Science test level/KS3 TA	N/A	0.32	0.64	0.63	0.54
	KS2 Average test level/KS3 TA	N/A	0.42	0.73	0.75	0.66

There are considerable fluctuations in the R^2 values for History (Table 4.12) and it is difficult, therefore, to make any general statement about the analysis. Key Stage 2 data clearly accounts for a fairly small proportion of the variation in Key Stage 3 data. The figures for CAT data are slightly better, but hardly impressive.

TABLE 4.12

REGRESSION (R^2) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN HISTORY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
History	CAT Verbal/KS3 TA	0.59	0.33	0.63	0.65	0.49
	CAT Quantitative/KS3 TA	0.30	0.31	0.51	0.52	0.33
	CAT Non-verbal/KS3 TA	0.31	0.20	0.41	0.49	0.35
	CAT Average/KS3 TA	0.49	0.35	0.64	0.58	0.49
	KS2 English test level/KS3 TA	N/A	0.12	0.43	0.44	0.32
	KS2 Maths test level/KS3 TA	N/A	0.16	0.40	0.37	0.32
	KS2 Science test level/KS3 TA	N/A	0.10	0.41	0.40	0.29
	KS2 Average test level/KS3 TA	N/A	0.18	0.54	0.56	0.44

In Geography, the R-values (Table 4.13) are clearly better for CAT data than for Key Stage 2 data. Whilst there are fluctuations, year on year, they do not generally seem to be as great as some of those found in the History data. The majority of these values are below 0.7, some of them substantially below.

TABLE 4.13

REGRESSION (R) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN GEOGRAPHY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
Geography	CAT Verbal/KS3 TA	0.77	0.61	0.71	0.75	0.62
	CAT Quantitative/KS3 TA	0.61	0.57	0.66	0.66	0.52
	CAT Non-verbal/KS3 TA	0.59	0.51	0.57	0.67	0.56
	CAT Average/KS3 TA	0.73	0.62	0.73	0.69	0.64
	KS2 English test level/KS3 TA	N/A	0.35	0.59	0.58	0.55
	KS2 Maths test level/KS3 TA	N/A	0.40	0.64	0.57	0.54
	KS2 Science test level/KS3 TA	N/A	0.35	0.65	0.58	0.48
	KS2 Average test level/KS3 TA	N/A	0.45	0.71	0.68	0.62

If we use the R^2 values for Geography (Table 4.14) we see that the CAT data suggests that some 50% to 60% of the variation in Key Stage 3 data cannot be accounted for using this form of baseline data. For Key Stage 2 data the figures are much higher, suggesting that some 70% to 80% of the variation cannot be explained using this data

TABLE 4.14

REGRESSION (R^2) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN GEOGRAPHY, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
Geography	CAT Verbal/KS3 TA	0.60	0.37	0.51	0.57	0.38
	CAT Quantitative/KS3 TA	0.37	0.32	0.44	0.44	0.27
	CAT Non-verbal/KS3 TA	0.35	0.26	0.32	0.45	0.31
	CAT Average/KS3 TA	0.54	0.39	0.53	0.47	0.41
	KS2 English test level/KS3 TA	N/A	0.12	0.35	0.34	0.30
	KS2 Maths test level/KS3 TA	N/A	0.16	0.41	0.32	0.29
	KS2 Science test level/KS3 TA	N/A	0.12	0.42	0.34	0.23
	KS2 Average test level/KS3 TA	N/A	0.20	0.51	0.46	0.38

Modern Foreign Languages again reveals higher R-values for CAT data than for Key Stage 2 data but there are also significant fluctuations in some year on year figures (Table 4.15).

TABLE 4.15

REGRESSION (R) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN MODERN FOREIGN LANGUAGES, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
MFL	CAT Verbal/KS3 TA	0.71	0.60	0.56	0.77	0.69
	CAT Quantitative/KS3 TA	0.62	0.62	0.63	0.66	0.56
	CAT Non-verbal/KS3 TA	0.56	0.57	0.46	0.66	0.69
	CAT Average/KS3 TA	0.71	0.66	0.62	0.70	0.73
	KS2 English test level/KS3 TA	N/A	0.42	0.45	0.62	0.57
	KS2 Maths test level/KS3 TA	N/A	0.47	0.46	0.61	0.55
	KS2 Science test level/KS3 TA	N/A	0.28	0.47	0.61	0.54
	KS2 Average test level/KS3 TA	N/A	0.48	0.52	0.72	0.66

The R^2-values for Modern Foreign Languages (Table 4.16) show that for CAT data, typically some 50% of variation in Key Stage 3 data cannot be accounted for by using this baseline data. The figure for Key Stage 2 data is at least 70% typically.

TABLE 4.16

REGRESSION (R^2) OF CAT DATA AND KEY STAGE 2 DATA WITH KEY STAGE 3 TEACHER ASSESSMENT DATA IN MODERN FOREIGN LANGUAGES, INTAKE YEARS 1994-1998.

Subject	Regression	1994	1995	1996	1997	1998
MFL	CAT Verbal/KS3 TA	0.50	0.36	0.31	0.60	0.48
	CAT Quantitative/KS3 TA	0.38	0.38	0.40	0.43	0.31
	CAT Non-verbal/KS3 TA	0.31	0.32	0.21	0.44	0.48
	CAT Average/KS3 TA	0.50	0.44	0.38	0.49	0.54
	KS2 English test level/KS3 TA	N/A	0.18	0.20	0.38	0.33
	KS2 Maths test level/KS3 TA	N/A	0.22	0.21	0.37	0.30
	KS2 Science test level/KS3 TA	N/A	0.08	0.22	0.37	0.29
	KS2 Average test level/KS3 TA	N/A	0.23	0.27	0.52	0.44

The other three National Curriculum subjects, Art, Music and Physical Education only started using numerical levels at Key Stage 3 in the summer of 2001 (1998 intake). It was therefore considered to be of little value to analyse only one year's worth of data for these subjects. Anecdotal evidence from teachers working in these areas, though, suggests that the regression values would be likely to be quite poor since either of the two forms of baseline data used for this study are unlikely to represent any kind of meaningful baseline measure for ability in these subjects.

Given the variation we have seen in the data for English, Mathematics and Science, it is interesting to look at the overall effects of combining these data sets into a Key Stage 3 average test level. The figures are somewhat surprising, as can be seen in Table 4.17. In general the highest coefficients are between the CAT average score and the Key Stage 3 average test level, though the CAT Verbal score also produces fairly high values. What is also apparent, however, is that year on year fluctuations are apparent in the data, once again, and that there is a general downward trend in these figures over

the five-year period. One possible explanation as to why these values are so high is that they are unduly distorted by the high values which we saw for Mathematics. Nevertheless the CAT average score would appear, on the basis of this evidence, to be a strong candidate for use as baseline data in evaluating and predicting the *average* Key Stage 3 test level in English, Mathematics and Science, though the overall trend towards a deterioration in the values for the CAT average score suggests grounds for a certain measure of caution, and also points to the need for further research in this area.

TABLE 4.17

REGRESSION (R) OF CAT DATA AND KEY STAGE 3 AVERAGE TEST LEVEL, INTAKE YEARS 1994-1998

Regression	**1994**	**1995**	**1996**	**1997**	**1998**
CAT Verbal/KS3 Average Test level	0.88	0.81	0.85	0.87	0.81
CAT Quantitative/KS3 Average Test level	0.79	0.75	0.79	0.82	0.69
CAT Non-Verbal/KS3 Average Test level	0.75	0.73	0.71	0.77	0.71
CAT Average/KS3 Average Test level	0.90	0.86	0.88	0.82	0.83

Use of the R^2-values to investigate these relationships further (Table 4.18) again underlines the potential value of the CAT average score for evaluating and predicting results at Key Stage 3 in English, Mathematics and Science since it explains some 70% to 80% of the variation in Key Stage 3 test data. However, the overall trend towards a deterioration in these values does, once again, suggest grounds for caution.

TABLE 4.18

REGRESSION (R^2) OF CAT DATA AND KEY STAGE 3 AVERAGE TEST LEVEL, INTAKE YEARS 1994-1998

Regression	**1994**	**1995**	**1996**	**1997**	**1998**
CAT Verbal/KS3 Average Test level	0.77	0.66	0.72	0.76	0.65
CAT Quantitative/KS3 Average Test level	0.63	0.56	0.63	0.67	0.48
CAT Non-Verbal/KS3 Average Test level	0.56	0.54	0.50	0.60	0.50
CAT Average/KS3 Average Test level	0.81	0.74	0.77	0.68	0.69

A further variation in outcomes is apparent if we aggregate Key Stage 2 test results, to form an average level, and examine the relationship between this figure and the Key Stage 3 average test level (Table 4.19). Once again these figures appear to be quite high by comparison with the figures for the individual subjects tested at Key Stages 2 and 3, though it is reasonable to assume that the figures are again distorted by those for Mathematics. These values are quite high, though the fluctuation over the four-year period for which data were available suggest that basing any assumptions on the figures at this stage would be premature.

TABLE 4.19

REGRESSION (R) OF KEY STAGE 2 AVERAGE TEST LEVEL WITH KEY STAGE 3 AVERAGE TEST LEVEL, INTAKE YEARS 1994-1998

Regression	**1994**	**1995**	**1996**	**1997**	**1998**
KS2 Average test level/KS3 Average test level	N/A	0.79	0.84	0.85	0.77

The R^2-values for this relationship (Table 4.20) show that some 30% to 40% of the variation in Key Stage 3 average test levels cannot be explained by the Key Stage 2 average test level, suggesting that it represents a questionable basis upon which to make assumptions or predictions.

TABLE 4.20

REGRESSION (R^2) OF KEY STAGE 2 AVERAGE TEST LEVEL WITH KEY STAGE 3 AVERAGE TEST LEVEL, INTAKE YEARS 1994-1998

Regression	**1994**	**1995**	**1996**	**1997**	**1998**
KS2 Average test level/KS3 Average test level	N/A	0.63	0.70	0.72	0.59

COMMENTARY ON THE VALUE-ADDED ANALYSES
(TABLES 4.1 – 4.20)

The full five-year data sets reflect, to a large extent, the findings of Moody (2001) in the initial two-year pilot study for this investigation. CAT data, in

particular the average scores, almost consistently produces higher coefficients with Key Stage 3 data, than does Key Stage 2 data. This relationship is much stronger for the three core subjects which are tested at Key Stage 3 than it is for the remaining teacher-assessed subjects. Overall, the coefficients for Mathematics suggest that the CAT average score provides a fairly reliable and valid basis upon which to make suppositions about value-added, and upon which to make predictions of future performance for the purposes of target-setting. The coefficients for English and Science are not as good, although they could be regarded as a reasonable basis for such procedures providing we take note of the possible instability of some of the relationships, and the possible deterioration in these relationships over time. All coefficients for the teacher-assessed subjects at Key Stage 3 are not only very poor, but fluctuate so wildly, that it must be concluded that there is neither any valid, or reliable, basis upon which any assumptions about value-added can be made, and that as a result, any attempt to predict future performance on this basis, with a view to setting targets, should be quite out of the question. These findings lend considerable support to the concerns raised by QCA (1998) about the consistency and reliability of teacher assessments at Key Stage 3.

The analyses clearly indicate that the relationships between the various sub-sets of data are not very stable over the five-year period. An obvious question is whether this is just a feature of the data for this particular case-study, or whether there is evidence of similar instability nationally. Useful evidence might be found hidden in data analysis carried out by the NFER. Commencing with the 1997 Key Stage 3 test result data, NFER (1998 – 2001) began to build up a database of test results for those students who had also taken the CAT. The first data-set was for 13,000 students and over the following years this data-set was increased so that it now includes students for the three most recent years in which Key Stage 3 tests have been taken. By 2001 this analysis was based on data from a nationally representative

sample of 98,751 students from 983 schools across 104 LEAs. Coincidentally, of course, these analyses also represent the five years across which this study was completed, so some useful comparisons can be made. NFER (2001) use "ordered logistic regression analysis" (page 1) to calculate the percentage chance of any individual student achieving a particular level at Key Stage 3 in English, Mathematics or Science. The CAT average score is used to make this calculation for Mathematics and Science, but the CAT Verbal test score is used to make the calculation for English because this has been found to have a stronger relationship with Key Stage 3 test results in English. If we calculate, from these data sets, the *average probability* of students achieving particular results at Key Stage 3, we see some interesting patterns emerging.

In English, for instance, (Table 4.21) there are significant fluctuations in the chances of a student of any given level of ability, as measured by the CAT, achieving any given level at Key Stage 3 in English.

TABLE 4.21

PROBABILITY OF ACHIEVING LEVELS 3-8 IN ENGLISH AT KEY STAGE 3 BASED ON NFER CHANCES TABLES FOR 1998-2001

CAT Verbal score	KS3 test year	Level 3 or less	Level 4	Level 5	Level 6	Level 7	Level 8
70-79	1998	21.50%	59.00%	16.50%	2.00%	1.00%	1.00%
	1999	48.50%	35.00%	14.50%	2.50%	0.00%	0.00%
	2000	46.00%	39.00%	13.50%	1.00%	0.00%	0.00%
	2001	42.50%	43.00%	13.50%	1.00%	0.00%	0.00%
80-89	1998	7.50%	49.50%	35.50%	7.00%	1.00%	1.00%
	1999	15.00%	37.50%	37.00%	9.50%	1.00%	0.00%
	2000	14.00%	41.00%	38.50%	5.50%	0.50%	0.00%
	2001	12.00%	44.00%	37.50%	5.00%	0.50%	0.00%
90-99	1998	2.50%	27.50%	48.00%	18.50%	3.50%	1.00%
	1999	4.00%	18.00%	46.00%	27.00%	4.50%	0.50%
	2000	4.00%	20.00%	54.50%	19.00%	2.50%	0.00%
	2001	3.00%	22.00%	55.00%	17.50%	2.50%	0.00%
100-109	1998	1.00%	11.00%	42.00%	36.00%	9.50%	1.00%
	1999	1.50%	6.00%	31.50%	46.00%	14.00%	1.50%
	2000	0.50%	7.00%	44.00%	39.50%	8.00%	0.50%
	2001	0.50%	7.50%	46.00%	37.00%	7.50%	0.50%
110-119	1998	1.00%	4.00%	23.50%	46.00%	24.00%	2.50%
	1999	0.00%	2.00%	15.00%	48.00%	31.00%	3.50%
	2000	0.00%	2.00%	24.00%	50.50%	20.50%	2.50%
	2001	0.00%	3.00%	27.50%	49.50%	18.00%	2.00%
120-129	1998	1.00%	1.50%	9.50%	36.50%	44.00%	8.50%
	1999	0.00%	0.50%	6.50%	35.50%	48.00%	9.50%
	2000	0.00%	1.00%	11.00%	45.00%	37.50%	6.00%
	2001	0.00%	1.50%	14.00%	48.00%	32.50%	4.50%

Further analysis of this data (Table 4.22) appears to show that for students in the 70-79 Verbal score range, they were considerably more likely to be awarded level 4, and level 3 or lower, in 2001, by comparison with 1997. For those with scores of 80-89 there was an increased chance of achieving level 3 or lower, at the lower end of this ability range, and a slightly increased chance of achieving level 5 at the higher end. For the 90-99 range the chances of achieving level 5 have risen by 7% between 1997 and 2001. In the average and slightly above average ability range (100-109), there is a

slightly improved chance of achieving level 5 (4%). At the 110-119 level the chances of achieving either levels 5 or 6 have risen by 3.5% - 4%. Students at the 120-129 level stood a higher chance of achieving a level 5 in 2001, and a much higher chance of achieving a level 6 (11.5%), but their likelihood of achieving a level 7 dropped by 11.5%. The chances of achieving level 8 also dropped by 4%.

What all this suggests is some quite complex and contradictory movements in the KS3 test data nationally. It would appear that students in the lowest ability range, as measured by the CAT Verbal test, are likely to achieve lower test levels at Key Stage 3 in English now than they were five years ago. In the middle ability range, we see some increased chances of achieving higher levels than in 1997, but paradoxically, levels 7 and 8 appear to be more difficult to achieve by the most able than they were five years ago. Given that the CAT is a standardised test battery, with high levels of reliability, as we have seen, this suggests that there is either a fairly fluid situation existing nationally in patterns of student learning, or there are complex patterns of instability in the National Curriculum tests for English at Key Stage 3. It is tempting to favour the latter explanation.

TABLE 4.22

PROBABILITY OF ACHIEVING LEVELS 3-8 IN ENGLISH AT KEY STAGE 3 IN 2001 BY COMPARISON WITH 1997

CAT Verbal score	**Level 3 or less**	**Level 4**	**Level 5**	**Level 6**	**Level 7**	**Level 8**
70-79	21.0% more likely	16.0% less likely	3.0% less likely	1.0% less likely	1.0% less likely	1.0% less likely
80-89	4.5% more likely	5.5% less likely	2.0% more likely	2.0% less likely	0.5% less likely	1.0% less likely
90-99	0.5% more likely	5.5% less likely	7.0% more likely	1.0% less likely	1.0% less likely	1.0% less likely
100-109	0.5% less likely	3.5% less likely	4.0% more likely	1.0% more likely	2.0% less likely	0.5% less likely
110-119	1.0% less likely	1.0% less likely	4.0% more likely	3.5% more likely	6.0% less likely	0.5% less likely
120-129	1.0% less likely	No change	4.5% more likely	11.5% more likely	11.5% less likely	4.0% less likely

There seems to be a different pattern of movement in the NFER data for Mathematics, as we can see in Table 4.23.

TABLE 4.23

PROBABILITY OF ACHIEVING LEVELS 3-8 IN MATHEMATICS AT KEY STAGE 3 BASED ON NFER CHANCES TABLES FOR 1998-2001

CAT Average Score	KS3 Test Year	Level 3 or less	Level 4	Level 5	Level 6	Level 7	Level 8
70-79	1998	64.00%	33.50%	2.50%	1.00%	1.00%	1.00%
	1999	68.50%	28.50%	2.50%	0.00%	0.00%	0.00%
	2000	67.00%	31.00%	2.00%	0.00%	0.00%	0.00%
	2001	64.50%	33.00%	2.50%	0.00%	0.00%	0.00%
80-89	1998	19.00%	58.50%	20.00%	3.00%	1.00%	1.00%
	1999	21.50%	55.50%	20.00%	2.50%	0.00%	0.00%
	2000	20.50%	56.00%	21.50%	2.00%	0.00%	0.00%
	2001	19.50%	55.00%	23.00%	3.00%	0.00%	0.00%
90-99	1998	2.50%	30.50%	44.00%	22.00%	2.00%	1.00%
	1999	3.00%	30.50%	45.50%	19.50%	1.50%	0.00%
	2000	2.00%	28.00%	48.00%	20.00%	1.50%	0.00%
	2001	2.00%	26.50%	46.50%	22.00%	2.50%	0.00%
100-109	1998	1.00%	4.50%	26.50%	53.50%	15.50%	1.00%
	1999	0.00%	4.50%	29.00%	50.50%	15.00%	0.50%
	2000	0.00%	3.50%	28.00%	50.00%	17.50%	0.50%
	2001	0.00%	3.50%	26.00%	46.00%	23.00%	1.00%
110-119	1998	1.00%	1.00%	4.00%	37.50%	53.00%	5.00%
	1999	0.00%	0.50%	4.50%	35.00%	52.50%	7.50%
	2000	0.00%	0.00%	3.50%	32.00%	55.00%	9.00%
	2001	0.00%	0.00%	3.50%	26.00%	57.00%	13.50%
120-129	1998	1.00%	1.00%	1.00%	7.00%	58.50%	33.50%
	1999	0.00%	0.00%	0.00%	5.50%	48.50%	46.00%
	2000	0.00%	0.00%	0.00%	4.50%	45.50%	50.50%
	2001	0.00%	0.00%	0.00%	3.00%	39.50%	56.50%

Further analysis of this data for Mathematics (Table 4.24) shows that the change in levels of achievement for those scoring 70-79 on the CAT average is negligible over the five year period. For those scoring 80-89 there appears to be a slight movement upwards from level 4 to 5. The same is true for those scoring 90-99. For those scoring 100-109 (average and slightly above average) there is clear movement from levels 6 to 7 (7.5%). At 110-119 the movement upwards appears to be fairly marked from level 6 to 7

(4.00%), and more notably from level 7 to 8 (8.5%). At 120-129 there is a very significant movement from level 7 to level 8.

What this suggests is that there has been no change in levels of achievement in Key Stage 3 Mathematics for those at the lowest ability levels. Middle ability students appear to be showing a modest trend towards upward movement, whilst at the higher ability levels there is a fairly marked upward movement towards levels 6 and 7, and for the most able a very marked increase from levels 7 to 8.

TABLE 4.24

PROBABILITY OF ACHIEVING LEVELS 3-8 IN MATHEMATICS AT KEY STAGE 3 IN 2001 BY COMPARISON WITH 1997

CAT Average score	**Level 3 or less**	**Level 4**	**Level 5**	**Level 6**	**Level 7**	**Level 8**
70-79	0.5% more likely	0.5% less likely	No change	1.0% less likely	1.0% less likely	1.0% less likely
80-89	0.5% more likely	3.5% less likely	3.0% more likely	No change	1.0% less likely	1.0% less likely
90-99	0.5% less likely	4.0% less likely	2.5% more likely	No change	0.5 % more likely	1.0% less likely
100-109	1.0% less likely	1.0% less likely	0.5% less likely	7.5% less likely	7.5% more likely	Ni change
110-119	1.0% less likely	1.0% less likely	0.5% less likely	11.5% less likely	4.0% more likely	8.5% more likely
120-129	1.0% less likely	1.0% less likely	1.0% less likely	4.0% less likely	19.0% less likely	23.0% more likely

Patterns of movement in Science data at Key Stage 3 are also very interesting as we can see in Table 4.25.

TABLE 4.25

PROBABILITY OF ACHIEVING LEVELS 3-8 IN SCIENCE AT KEY STAGE 3 BASED ON NFER CHANCES TABLES FOR 1998-2001

CAT Average score	KS3 Test Year	Level 3 or less	Level 4	Level 5	Level 6	Level 7	Level 8
70-79	1998	46.00%	47.00%	6.50%	1.00%	1.00%	1.00%
	1999	62.50%	33.00%	4.50%	0.50%	0.00%	0.00%
	2000	60.50%	36.50%	3.00%	0.00%	0.00%	0.00%
	2001	64.50%	31.00%	4.00%	0.50%	0.00%	0.00%
80-89	1998	13.00%	56.50%	27.00%	3.50%	1.00%	1.00%
	1999	24.00%	54.00%	19.50%	2.50%	0.00%	0.00%
	2000	22.00%	60.00%	16.50%	1.00%	0.00%	0.00%
	2001	27.00%	51.00%	19.50%	2.00%	0.00%	0.00%
90-99	1998	2.50%	19.00%	50.00%	17.50%	1.50%	1.00%
	1999	5.50%	36.00%	45.00%	12.00%	1.00%	0.00%
	2000	4.00%	41.50%	45.00%	8.50%	0.50%	0.00%
	2001	6.00%	35.00%	46.50%	12.00%	0.50%	0.00%
100-109	1998	1.00%	7.00%	38.00%	46.00%	8.50%	1.00%
	1999	0.50%	10.00%	42.00%	39.50%	7.50%	0.00%
	2000	0.50%	12.00%	49.00%	34.00%	4.50%	0.00%
	2001	1.00%	9.50%	43.50%	41.00%	5.00%	0.50%
110-119	1998	1.00%	1.50%	12.00%	51.50%	34.00%	1.00%
	1999	0.00%	1.50%	14.00%	47.00%	35.50%	1.50%
	2000	0.00%	1.50%	19.50%	51.00%	26.00%	1.50%
	2001	0.00%	1.00%	15.00%	54.50%	26.50%	3.00%
120-129	1998	1.00%	1.00%	2.50%	24.50%	68.00%	5.00%
	1999	0.00%	0.00%	1.50%	18.00%	64.00%	16.00%
	2000	0.00%	0.00%	3.50%	27.50%	56.00%	13.00%
	2001	0.00%	0.00%	2.00%	26.00%	51.00%	21.00%

Further analysis of this data for Science (Table 4.26) reveals that for the least able students with a CAT average score of 70-79 the movement is clearly downwards, particularly from level 4 to level 3 or lower. This also appears to be true for those with scores of 80-89. There appears to be downward movement for those with scores of 90-99, across levels 6, 5 and 4. In the 100-109 range there is downward movement across levels 7, 6 and 5, a pattern repeated for those at 110-119 (with a slight upward movement to level 8, presumably at the top end of this range). At 120-129 there is a

very marked movement from level 7 to level 8 (16%). This shows a different pattern again from either English or Mathematics.

TABLE 4.26

PROBABILITY OF ACHIEVING LEVELS 3-8 IN SCIENCE AT KEY STAGE 3 IN 2001 BY COMPARISON WITH 1997

CAT Average Score	Level 3 or less	Level 4	Level 5	Level 6	Level 7	Level 8
70-79	18.5% more likely	16.0% less likely	2.5% less likely	0.5% less likely	1.0% less likely	1.0% less likely
80-89	14.0% more likely	5.5% less likely	7.5% less likely	1.5% less likely	1.0% less likely	1.0% less likely
90-99	3.5% more likely	6.0% more likely	3.5% less likely	5.5% less likely	1.0% less likely	1.0% less likely
100-109	No change	2.5% more likely	5.5% more likely	5.0% less likely	3.5% less likely	0.5% less likely
110-119	1.0% less likely	0.5% less likely	3.0% more likely	3.0% more likely	7.5% less likely	2.0% more likely
120-129	1.0% less likely	1.0% less likely	0.5% less likely	1.5% more likely	17% less likely	16% more likely

What these analyses appear to show is complex movements at Key Stage 3 across all three subject areas that are tested, and this may go some way towards explaining the fluctuations, and possibly the deteriorating relationships, between CAT data and Key Stage 3 test results. This casts further doubt on both the validity and the reliability of value-added procedures, assumptions, or comparisons at Key Stage 3, and by implication therefore raises questions about the validity and reliability of target-setting procedures based on such data, particularly using linear regression. The only possible exception is for Mathematics where the scalable nature of the

knowledge and skills in this area may form the underlying reason why validity and reliability levels appear to be so much higher in this subject.

The weak figures found in the teacher-assessed subjects at Key Stage 3 could be caused by a number of factors. One obvious question, as we noted with Art, Music and Physical Education, is whether Key Stage 2 data, or CAT data, can be regarded as a valid baseline measure by which to make judgements about progress in these teacher-assessed subjects. In subjects like History and Geography it is difficult to see why the kind of core academic skills represented by Key Stage 2 tests, or the standardized measures of intelligence as represented by the CAT, should not be a reasonably reliable basis upon which to base such judgements. On the other hand, it is possible that subjects such as Design Technology, Information and Communications Technology and Modern Foreign Languages are calling upon very different kinds of skills and abilities from those measured by either Key Stage 2 tests, or the CAT. However, whilst there are clearly debates surrounding the concept of intelligence, it is a little difficult to see how a national age-standardized test of intelligence could give rise to such poor relationships with these subjects at Key Stage 3.

As we have seen, QCA (1998) found that there were significant problems for teachers in making assessments in non-core subjects at Key Stage 3 because of the different ways in which the criteria for each level could be interpreted. Teachers reported their own feelings that their judgements were not achieving consistency, mentioning in particular the problem of borderline students. Effectively having to assign any student to only one of four categories is problematic, especially when there are no borderline criteria. Furthermore, as we noted, as a result of this, any teacher stands a twenty-five per-cent chance of classifying a student at the "wrong" level, inasmuch as it is possible to identify, with any degree of certainty, what is right or wrong in this context. Most teachers reported that they did not believe that "the 'best-fit' approach" (p.10) worked well in terms of

consistency, especially for borderline cases. Some appeared to be using an average over the academic year to arrive at the final level, some were looking for key aspects of the level descriptors having been met, whilst others were using a best-fit model. As the report recommended "it is important that further work be carried out to investigate exactly how teachers define 'best-fit'" (p.14) since some teachers were using quantitative data, others were using the "hurdle" approach and some were just using intuition. "Progression was found to be particularly problematic because of the difficulties in comparing the same levels across Key Stages, and thus, ages" (p.16). This meant that teachers felt that they could not regard the 8-level National Curriculum scale as being on a continuum. Whilst many teachers held internal moderation meetings to try to standardize their assessments, there was little evidence of moderation across schools and the report recommended that "cross-school agreement trialling" was the "next most important training need" (p.84). To date this has not happened. However, even if this recommendation had been implemented it is tempting to wonder how reliable such a local procedure would be. Surely what is needed is a form of national sample moderation such as we use for GCSE coursework, where there is every reason for the teacher to get the judgement right to avoid all the students in the school being penalised. It could be argued of course, on the other hand, that this would not remove the problem of inadequate level descriptions. Given the inadequacy of such a limited number of levels, and given the need for more finely-differentiated data for the purposes of making value-added judgements, perhaps it is time to abandon levels and start awarding percentages.

In an attempt to ascertain whether using more finely differentiated data at Key Stage 3, would improve value-added relationships, regressions were calculated for the CAT scores and the actual *marks* awarded to students who took their Key Stage 3 English tests in 2001 (1998 intake). The results can be seen in Table 4.27. What this analysis suggests is that using more finely differentiated data does not necessarily strengthen the nature of the

relationship between the two sets of data, using linear regression techniques, and therefore the validity of this method must also be in some doubt. The strongest relationship is between the CAT Verbal test data and the Key Stage 3 English mark, though at 0.68 this is fairly weak. However, since this analysis has only been carried out on one data set, there is clearly room for further investigation of this issue. There are a number of problems, however, associated with undertaking this kind of analysis in Mathematics and Science because, unlike in English, there are different levels of entry with different mark schemes which are not easily equated.

TABLE 4.27

REGRESSION (R) OF CAT SCORES AND MARKS ON THE KEY STAGE 3 TEST IN ENGLISH IN 2001 (1998 INTAKE)

Regression	
CAT Average/KS3 English mark	0.61
CAT Non-verbal/KS3 English mark	0.46
CAT Quantitative/KS3 English mark	0.46
CAT Verbal/KS3 English mark	0.68

TARGETS AND TARGET-SETTING DATA: ANALYSIS AND DISCUSSION

Targets were first set in the school in which this study was conducted in September of 1999. Tables 4.29 – 4.52 represent the results of the questionnaires to students, staff and parents and reveal a number of interesting insights into how target-setting procedures affected performance, and how students, staff and parents perceived these procedures. (The questionnaires used to gather this data can be found in Appendices 1-4). It is important to remember that these targets were "high-challenge" (see Chapter 1) and were deliberately designed to try to raise levels of performance by setting the target at one level above what the national statistics suggested that the majority of students, at any given level of prior ability, as measured by the CAT average score, would be likely to achieve.

EXTENT TO WHICH STUDENTS ACHIEVED THEIR TARGETS, 2000-2002

Since the school set so-called high-challenge targets (see Chapter 1) for each student, based on their average CAT score, and since the target was one National Curriculum level above what they would normally be expected to achieve according to the NFER chances tables, the percentages of students achieving their high challenge target, or above, between 2000 and 2002 are surprisingly high in many subjects, as we can see in Table 4.28. There is a marked disparity in the results for Science, by comparison with those for English and Mathematics, although there is strong evidence of rising standards in all three subjects. In the non-core subjects, again, there is an interesting range of differences between subject areas, and some interesting year-on-year changes, such as the 13% drop between 2000 and 2001 in Geography and the 10% rise over the same two years in ICT. This raises the question as to whether the differences can be explained in terms of the problems already discussed with regard to teacher assessment of these subjects, or whether other factors account for these differences. Whatever the reason may be, the figures raise some interesting questions with regard to target-setting in the non-core subjects.

TABLE 4.28

PERCENTAGES OF STUDENTS IN THE 1997-1999 INTAKE YEARS ACHIEVING THEIR HIGH CHALLENGE TARGET, OR ABOVE, IN 2000 – 2002.

Subject	2000	2001	2002
English	64%	67%	86%
Mathematics	63%	68%	67%
Science	44%	50%	60%
Modern Languages	74%	60%	53%
ICT	49%	59%	44%
DT	77%	77%	72%
History	59%	56%	66%
Geography	62%	49%	60%
Art	N/A	97%	63%
Music	N/A	75%	54%
PE	N/A	46%	45%

It is difficult to estimate what we can safely conclude from Table 4.28 about the reliability of target-setting procedures in use. In Mathematics and DT the figures are stable over the three year period. In some subjects there is a steady upward trend (English and Science) whilst in many others there are unpredictable fluctuations year on year. Certainly using targets in subjects where these kinds of fluctuations are evident would appear to be extremely problematic.

The consequential validity of using targets is clearly a critical issue, because of the feedback effects and expectations which they give rise to. The questionnaires to staff, students, and parents, were designed, in part, to explore this issue.

STAFF QUESTIONNAIRE: ANALYSIS AND COMMENTARY

The total number of respondents to this questionnaire was 36, representing approximately half of the staff employed in the school. The percentages quoted in each of the tables should be interpreted in this context.

The figures in Table 4.29 reveal some very interesting patterns and it is instructive to compare these with the analysis of the staff interviews in the following chapter.

Whilst 47% of staff think that the mechanisms for setting targets are good, or very good, it is tempting to wonder why 14% don't know. Does this represent a refusal to engage with the targets debate on ideological grounds? Does it reflect the fact that despite the efforts of the school to keep staff informed about how targets are calculated, staff take little notice of this information? Does it mean that those staff do not understand the mechanisms? It is interesting to note that this question produced the highest figure in the "poor" column, 11%.

Given that the school decided to set a target-average for all subjects, it is interesting to note that 56% of staff think that the target is good, or very good, for their subject area. This does imply, though, a fair measure of concern and comments added to the questionnaire reveal concerns in particular subject areas, as we will see.

Staff estimates of how well students understand the "high challenge" aspect of targets are not as strong as the estimates by students themselves, as we will see. 50% of staff rated student understanding as being good, or very good. 55% of staff think that the extent to which targets help to improve learning is good, or very good. The fact that 6% don't know again raises questions. Why don't they know? Is it an unwillingness to think about the issue? Is it ideological?

The figures on the extent to which staff see targets as helping to improve teaching are reasonably high, given the concerns which abound with regard to target-setting. 53% saw targets as being good, or very good, with regard to improving teaching. Attitudes towards targets by students were seen by 55% of staff as being good or very good, although once again we see 11% of staff failing to express a judgement. Does this reflect the ambivalence which we will see later in the staff interviews?

In terms of parental attitudes, 47% of staff believed that parents saw targets as being good, or very good, but a very surprising 20% said they did not know what parents thought. Given that targets had been in use for nearly two years when this survey was conducted, and given that staff had been encouraged to discuss student progress at parents' evenings in terms of targets, this figure is difficult to explain. It suggests that staff might not be using the targets to discuss progress.

Staff were offered the opportunity to add comments to the end of the questionnaire and some of these were very instructive. Parents were seen by some staff as failing to understand targets, whilst others saw them as having very varied attitudes. Other parents were seen as being very well-informed about the issue. Students were seen by one teacher as being "mesmerized" by targets, whilst others took the view that students saw targets as being something that they were entitled to, and then pressurized teachers to achieve this for them. As we will see, this theme surfaced again during the staff interviews. There was a perception that targets were quite good for average and above-average students but that they did not work so well for below-average students, a theme which again emerged in the staff interviews, but which is not necessarily supported by the findings of the interviews with less able students. The extent to which targets were seen to improve learning was seen as depending very much on the individual pupil by some staff. Particular concern was expressed about the suitability of

targets in PE since the baseline data used to set targets was seen as having no relevance to the kinds of skills and knowledge which was needed to succeed in this subject. One might have expected similar concerns to be expressed for Art and Music, but these were not apparent.

TABLE 4.29

RESULTS OF QUESTIONNAIRE TO STAFF EVALUATING TARGET-SETTING PROCEDURES (MARCH 2001)

	Poor	Fair	Good	Very good	Excellent	Comment only	Don't know
Mechanism for deciding targets	11%	28%	36%	11%	0%	0%	14%
How realistic are the targets for your subject area?	5%	31%	45%	11%	0%	8%	0%
How well do students understand the "high challenge" aspect of targets?	8%	34%	42%	8%	0%	3%	5%
Extent to which targets help to improve learning	3%	31%	47%	8%	0%	5%	6%
Extent to which targets help to improve teaching	8%	39%	39%	14%	0%	0%	0%
Attitudes towards targets by students	5%	20%	47%	8%	0%	9%	11%
Attitudes towards targets by parents	3%	25%	33%	14%	0%	5%	20%

These figures reveal some very interesting patterns and it is instructive to compare these with the analysis of the staff interviews in the following chapter.

INITIAL TARGET EVALUATION QUESTIONNAIRES WITH YEAR 7, 1999-2001: ANALYSIS OF INDIVIDUAL QUESTION RESPONSES

The population sizes for these evaluations were as follows:

1999 N = 160 2000 N = 176 2001 N = 173

The percentages quoted in each of the tables should therefore be interpreted in the context of this information.

These surveys were taken a few months after students had been given their targets. The targets are given to tutors, subject staff, and to parents, and students have to record their target in their personal planners. Students are expected to know their targets and refer to them in the classroom when talking to staff. Table 4.30 reveals that whilst there is a steady improvement in the figures between 1999 and 2001, it is remarkable that one in ten students still did not know what their target was in 2001 when the procedure was in its third year of operation.

TABLE 4.30

KNOWLEDGE OF TARGETS IN YEAR 7, 1999-2001

	1999	2000	2001
Correct	60%	80%	91%
Incorrect	40%	20%	9%

Student feelings when they were first given their targets reveal some very contradictory trends as we can see in Table 4.31. Overall there seem to be fewer students who are very anxious, or anxious by 2001, but on the other hand there seems to be a reduction in the percentage of students who are not anxious at all. Do these patterns reflect student attitudes, parental attitudes, or the attitudes of the school, teachers, or tutors?

TABLE 4.31

FEELINGS WHEN FIRST GIVEN THE TARGET IN YEAR 7, 1999-2001

	1999	2000	2001
Very anxious	4%	6%	2%
Anxious	18%	24%	12%
A little bit anxious	48%	50%	61%
Not anxious at all	30%	20%	25%

The figures showing how well students think that they understood the target, (Table 4.32) when they were first given it in Year 7 remain more or less constant across the three year period with nearly two-thirds consistently saying that they understood and a further third saying that they felt that they clearly understood. A small proportion claimed not to be able to understand what the target represented. This is very encouraging in terms of the possible validity of the procedures.

TABLE 4.32

UNDERSTANDING OF THE TARGET WHEN FIRST GIVEN, YEAR 7, 1999-2001

	1999	2000	2001
Didn't understand at all	5%	6%	3%
I think I understood	66%	63%	64%
I clearly understood	29%	31%	33%

There is a very positive trend in terms of the way that students feel about their targets when they are first given to them, as we can see in Table 4.33. Some 80%-90% of students seem to perceive their high-challenge target as being about right for what they perceive as their ability level. This is interesting given that the concept of the high-challenge target being one level above what they might be expected to get is explained to students when they are first given their targets, and that tutors, and staff, in assemblies, reinforce this concept from time to time. Whilst it could be argued that students are merely attempting to protect their own self esteem by choosing "about right", it is interesting that there has been a significant drop in the proportion of students who see the target as being too low, and an increase in the proportion seeing the target as too high. Does this reflect a greater degree of realism on the part of students, or their parents?

TABLE 4.33

FEELINGS ABOUT ABILITY TO MEET THE TARGET WHEN FIRST GIVEN, YEAR 7, 1999-2001

	1999	2000	2001
Too low	9%	10%	4%
About right	87%	84%	90%
Too high	2%	3%	6%
No answer	2%	3%	0%

Despite considerable efforts to persuade students and tutors to discuss progress towards their targets with their personal tutors, the figures in Table 4.34 are lamentable. No more than 6% of students ever seem to have engaged in such a conversation. However a particular structural problem for the school has been the fact that the school is split across three sites and "substitute" tutors frequently have to cover the pastoral roles of their colleagues. It is often very difficult, therefore, for tutors to build up the kind of relationships with students in their tutor-groups which might facilitate this kind of conversation. On the other hand it seems remarkable that so few students have ever talked to their tutors, despite repeated requests to staff to implement such discussions.

TABLE 4.34

STUDENT CONVERSATIONS WITH THEIR PERSONAL TUTORS ABOUT THEIR TARGETS, YEAR 7, 1999-2001

	1999	2000	2001
Yes	4%	5%	6%
No	96%	95%	94%

By contrast, conversations between parents and students about their targets appear to be going on with considerably more frequency (some 60%-70%), as we can see in Table 4.35.

TABLE 4.35

STUDENT CONVERSATIONS WITH THEIR PARENTS ABOUT THEIR TARGETS, YEAR 7, 1999-2001

	1999	2000	2001
Yes	60%	62%	72%
No	40%	38%	28%

The figures for the percentage of students who think they will achieve their target in most of their subjects remains reasonably constant over the period, (see Table 4.36), at around 60%, but the proportion who think they will achieve their target in *all* of their subjects has nearly halved over the three years, from 11% to 6%. The proportion of those who see themselves as being unlikely to achieve their targets in any subjects remains relatively constant. These figures offer an interesting insight into student perceptions of their own ability, projected forward by three years. Given that the only evidential basis that they have for these perceptions is their performance at Key Stages 1 and 2, it is interesting to speculate as to how this perceptual mechanism works. Do these figures represent aspirations rather than perceptions?

TABLE 4.36

PERCEIVED LIKELIHOOD OF REACHING TARGET, YEAR 7, 1999-2001

	1999	2000	2001
I don't think I will reach my targets in any of my subjects	3%	1%	3%
I think I might reach my targets in a few of my subjects	27%	36%	31%
I think I might reach my targets in most of my subjects	59%	53%	60%
I think I might reach my targets in all of my subjects	11%	10%	6%

The general stability of these figures showing the overall reaction of students to the idea of having targets is interesting over a three year period (Table 4.37). Some 80%-90% of students see targets as being likely to make them work harder in some, or all of their subjects. The only real change is in those who see targets as being unlikely to make any difference to the way in which they will work, with a drop from 16% to 11%. Does this reflect a greater level of commitment by students or parents? As National Curriculum tests become ever-more "high stakes", is this changing attitudes?

TABLE 4.37

OVERALL REACTION TO THE IDEA OF HAVING TARGETS, YEAR 7, 1999-2001

	1999	2000	2001
I don't think targets will make any difference to the way I work	16%	12%	11%
I think targets might encourage me to work harder in some subjects	52%	59%	55%
I think targets will encourage me to work hard in all of my subjects	32%	29%	34%

COMMENTARY ON ANALYSIS OF YEAR 7 QUESTIONNAIRE (TABLES 4.30-4.37)

The analysis suggests that there are some interesting trends developing when students are first given their targets in Year 7, which form an interesting counterpoint to the perceptions of staff. The apparently reducing level of anxiety is somewhat surprising, as is the general support for the idea of targets. The evidence with regard to student perceptions of their own ability in relation to their targets opens up some interesting avenues for

further research. It is clear that parents are becoming more involved in the performance of their daughters in relation to their targets, even at this early stage in their secondary education, though as we will see in the student interviews in Chapter 5, questions arise as to how seriously students really take targets, once the initial flurry of interest wears off. The fact that some 10% did not know their target in Year 7, even when the scheme was in its third year of operation raises some interesting questions about the validity and reliability of the procedures in practice. As we will see, the analysis of questionnaires for the end of Year 9 suggests that this 10% figure remains relatively constant across Key Stage 3 for the three years in which the target-setting procedures were evaluated. An obvious question is whether this is a problem with low ability students or whether it extends across the ability range. Again, there is room here for further research. As we will see in the staff interviews, there is evidence that teachers treat targets for low ability students with considerable caution. Maybe a lack of reference to targets with these students encourages them to perceive targets as being unimportant.

FINAL TARGET EVALUATION QUESTIONNAIRES WITH YEAR 9 STUDENTS, 2000-2002: ANALYSIS OF INDIVIDUAL QUESTION RESPONSES

The population sizes for these questionnaires were as follows:

2000 N = 122 2001 N = 150 2002 N = 157

The percentages quoted in the tables should be interpreted in the context of this information.

These questionnaires were given to students after they had received their Key Stage 3 test results and teacher assessments. Students were therefore able to evaluate their targets in the light of this information. The figures in Table 4.38, showing the extent to which students actually know what their

target is at the end of Year 9, suggests an encouraging improvement over the three year period.

TABLE 4.38

KNOWLEDGE OF TARGETS, YEAR 9, 2000-2002

	2000	2001	2002
Correct	80%	92%	98%
Incorrect	20%	8%	2%

It is clear that the overwhelming majority of students think that the target level is about right (Table 4.39). This is interesting given that they are looking back over three years of work across all Key Stage 3 subjects, and that the target is an average across all these subjects. It is also interesting that a diminishing minority think that the target was too low in retrospect.

TABLE 4.39

STUDENT VIEW OF THE TARGET LEVEL, YEAR 9, 2000-2002

	2000	2001	2002
Too low	16%	13%	9%
About right	79%	83%	83%
Too high	3%	4%	8%
No answer	2%	0%	0%

Half to two thirds of the students see an average target as being suitable for most subjects, as we can see in Table 4.40. On the other hand it should be a matter of concern that a quarter to a third of students see the target as only being suitable for a few subjects. The fact that such a tiny percentage see the target as being unsuitable for any subjects is clearly very positive.

TABLE 4.40

STUDENT VIEW OF THE APPROPRIATENESS OF THE TARGET ACROSS SUBJECTS, YEAR 9, 2000-2002

	2000	2001	2002
Not right for any subjects	3%	4%	2%
Only right for a few subjects	35%	31%	25%
About right for most subjects	55%	57%	68%
Right for all subjects	6%	8%	5%

About half of students see targets as having effects in at least some subjects, and a varying percentage see them as having an effect in all subjects (Table 4.41). On the other hand a falling percentage of students see targets as making no difference to the amount of effort they make because they feel that they always do their best anyway. It would represent an interesting piece of further research to see to what extent these perceptions are shared by the staff who teach them.

TABLE 4.41

EFFECTS ON STUDENT EFFORT OF HAVING TARGETS, YEAR 9, 2000-2002

	2000	2001	2002
No effects in any subject	17%	10%	22%
Effects in some subjects	46%	48%	54%
Effects in all subjects	12%	19%	16%
Always work hard, so need for targets	25%	23%	8%

The figures for students talking to their personal tutors about their targets remain extremely poor, even at the end of Year 9, with some 70%, on average, reporting that they have never talked to their tutors (Table 4.42).

TABLE 4.42

STUDENTS TALKING TO PERSONAL TUTORS ABOUT THEIR TARGETS, YEAR 9, 2000-2002

	2000	2001	2002
Never	73%	72%	67%
Once or twice	26%	26%	31%
Reasonably frequently	1%	2%	1%
Very frequently	0%	0%	1%

It is difficult not to conclude that targets play little part in classroom discussions about progress either (Table 4.43), with some two-thirds of students reporting that conversations with classroom teachers had only taken place once or twice, and the rest reporting that such conversations never took place. This opens up interesting questions about what effects targets are really having at classroom level.

TABLE 4.43

STUDENTS TALKING TO SUBJECT TEACHERS ABOUT THEIR TARGETS, YEAR 9, 2000-2002

	2000	2001	2002
Never	27%	34%	24%
Once or twice	65%	56%	63%
Reasonably frequently	7%	10%	10%
Very frequently	1%	0%	3%

On the other hand student attitudes towards targets seem to be very positive (Table 4.44). These figures suggest that around 90% of students see targets as being either sometimes helpful, or very helpful. This is a very positive picture, despite the problems suggested by some of the earlier analyses.

TABLE 4.44

STUDENT ATTITUDES TOWARDS TARGETS, YEAR 9, 2000-2002

	2000	2001	2002
No use at all	11%	12%	9%
Sometimes helpful	66%	71%	76%
Very helpful	23%	17%	15%

COMMENTARY ON THE YEAR 9 QUESTIONNAIRE ANALYSIS (TABLES 4.38-4.44)

The problem of the 10% who do not know what their target is, even after three years, was referred to earlier. This represents a significant problem in terms of the supposed validity and reliability of target-setting procedures. Student perceptions of targets again opens up questions as to what kind of effects targets are having on these perceptions. Are targets causing students to classify themselves in ability terms which they then see subsequently as being approximately correct, or do students genuinely see the concept of a high challenge target as having raised their level of performance? The concept of an average target across all subjects is clearly seen as being potentially more problematic, whilst effects on student motivation are variable. The evidence for effects on motivation in some subjects is clear, and as we will see in Chapter 5, this is supported by evidence from student interviews. Evidence that targets make no difference at all can also be found in the interviews with more able students in Chapter 5. The figures on subject teachers referring to targets in lessons are very poor and suggest that targets are playing a very limited role in improving learning on a day to day basis. This again raises questions about the validity and reliability of target-setting procedures. On the other hand, general student attitudes towards targets do seem to be very positive and suggest that may be important positive consequential effects of using targets,

PARENTAL ATTITUDES QUESTIONNAIRE: ANALYSIS OF INDIVIDUAL QUESTION RESPONSES

The questionnaires were completed in the summer of 2001.The total number of respondents to the questionnaire was 135. This represents approximately 25% of the total parental population. The percentages quoted in the following tables should be interpreted in the context of this information.

These questionnaires were given to parents in the summer of 2001 when targets had been in operation for two years. The figures on parental familiarity with their daughter's target are clearly very positive, (Table 4.45) though the fact that some one in ten parents, who took the trouble to reply to this questionnaire, did not know their child's target is interesting.

TABLE 4.45

LEVEL OF FAMILIARITY WITH CHILD'S TARGET

Not familiar at all	11%
I think I know what the target is	31%
I am quite clear what the target is	58%

The school actually goes to some lengths to ensure that parents are both informed of the target at the beginning of each Key Stage, and are given information about how the target is calculated. The concept of the high challenge target is also explained. It is surprising, therefore, as we can see in Table 4.46, that some 22% of parents see the school's success in this area as being poor, or very poor. It is tempting to wonder whether parents who gave this response read the information sent to them. On the other hand, perhaps the school needs to find out precisely what it is that these parents think is poor about this aspect of the school's performance.

TABLE 4.46

SUCCESS OF SCHOOL IN EXPLAINING HOW THE TARGET WAS CALCULATED

Very poor	2%
Poor	20%
Satisfactory	29%
Generally good	37%
Very good	12%

Table 4.47 suggests that the overwhelming majority see the target as about right, but why nearly 10% gave no answer to this question is curious. Is this the 10% who do not know what the target is? It is interesting that 8% see the target as too low.

TABLE 4.47

FEELINGS ABOUT THE HIGH CHALLENGE TARGET

Too low	8%
About right	83%
Too high	0%
No answer	9%

Parental attitudes towards the concept of one target across all subject areas is quite positive, showing that some 65% of parents see the target as being right for most, or all, of the subjects (Table 4.48). Once again though, it is tempting to wonder why 13% gave no answer to this question.

TABLE 4.48

FEELINGS ABOUT HAVING ONE TARGET FOR ALL SUBJECTS

Not right for any subjects	0%
Only right for a few subjects	22%
About right for most subjects	51%
Appropriate for all subjects	14%
No answer	13%

Overall 46% of parents thought that their child worked harder in some, or all, subjects as a result of having a target (Table 4.49).The figure of just over

a third seeing the target as making no difference is actually higher than that found in the student questionnaire. Once again the reason for 10% giving no answer is a mystery.

TABLE 4.49

PERCEPTIONS ABOUT THE EFFECTS OF THE TARGET ON CHILD'S EFFORT

Didn't work harder in any subject	10%
Worked harder in some subjects	27%
Worked harder in all subjects	19%
Doesn't need a target to work hard	34%
No answer	10%

48% of parents clearly talk to their child about progress towards their target, either reasonably, or very, frequently (Table 4.50). Taken with the 33% who talk with the child now and again, this is fairly encouraging. On the other hand the fact that 16% never talk to the child potentially reveals quite a lot about parental interest.

TABLE 4.50

PARENTAL CONVERSATIONS WITH THE STUDENT ABOUT PROGRESS TOWARDS THE TARGET

Never	16%
Once or twice	33%
Reasonably frequently	36%
Very frequently	12%
No answer	3%

88% of parents see targets as being sometimes, or always, helpful (Table 4.51). The one in ten who have not given an answer have reappeared in this question.

TABLE 4.51

PARENTAL FEELINGS ABOUT THE IDEA OF TARGETS

Targets are not helpful	3%
Targets are sometimes helpful	63%
Targets are always helpful	25%
No answer	9%

91% of parents see the reports as being reasonably clear, or very informative, with regard to information about progress towards the targets (Table 4.52), an encouraging figure.

TABLE 4.52

PARENTAL EVALUATION OF INFORMATION IN STUDENT REPORTS ABOUT PROGRESS TOWARDS THE TARGET

Reports are not very helpful	5%
Reports are reasonably clear	59%
Reports are very informative	32%
No answer	4%

PARENTAL QUESTIONNAIRE: COMMENTARY ON THE ANALYSIS (TABLES 4.45-4.52)

There seems to be a range of opinion in the questionnaire responses, but the overall pattern suggests a reasonably good level of parental support for targets, reporting of progress towards targets, and for the concept of using targets to try to raise performance. On a number of questions the percentage of parents not offering an answer is interesting. Was this just an administrative error on their part, did it represent a question that they found difficult to answer, or were they concerned about what it might reveal? There is a general sense of relaxed support for what the school is trying to do and this reflects the general attitudes detected in the student interviews, as we will see in the next chapter. This is in contrast to teacher concerns. Given that it is reasonable to suppose that a number of parents are likely to have performance targets in their own workplace, it is interesting that

parental attitudes seem to be quite relaxed about student targets. Is this because parents are shielding their children from the more corrosive pressures of performance targets? Or is it because they see it as the job of the school, and teachers, to deliver the target performance? The overall impression, therefore, is that the consequential effects of students having targets are very positive as far as parental attitudes are concerned.

SUMMARY

It seems fairly clear, from the analysis of the quantitative data on value-added procedures that there are a number of concerns with regard to the use of these procedures in English and Science, both in terms of the validity and reliability of these procedures. In Mathematics there are clearly much stronger grounds for assuming that levels of validity and reliability are considerably higher. It is clear in all three subjects that Key Stage 2 data is less valid and reliable as a baseline than CAT data. In non-core subjects levels of validity and reliability are so low as to make value-added procedures meaningless, not least because of the immense fluctuations in the data, year on year. The data on targets suggests that these procedures are producing raised standards of performance, and that there is good general support from students, staff and parents, for these procedures. On the other hand, the fact that some 10% of students do not appear to know what their target is, even at the end of Key Stage 3, raises some questions about how valid and reliable the procedure can be assumed to be in practice. As we will see in the following chapter, however, detailed interviews with staff and students reveal a different set of perspectives on both value-added and target-setting procedures.

CHAPTER 5

ATTITUDES AND PERCEPTIONS

ANALYSIS AND DISCUSSION OF QUALITATIVE DATA

Extended interviews were conducted with a number of staff and students, as detailed in Chapter 2. A series of question prompts was created based on issues arising from analysis of the numerical data (see Appendices 5 and 6). The interviews were semi-structured, using the questions as a basis for exploration of a range of attitudes and perceptions surrounding the issues of value-added and target-setting procedures.

INTERVIEWS WITH STAFF

BASELINE DATA, VALUE-ADDED AND TARGET-SETTING

The concept of baseline data, and its perceived validity with regard to making value-added calculations, predictions, and setting targets based on such procedures, produced an interesting range of reactions. All staff saw value-added procedures as being useful up to a point, but there were a range of concerns expressed about the concept. Typical of the positive reactions was from a Science teacher who commented:

"It's a useful benchmark..... a way of getting teachers to conceptualize the idea of *progress*....it's therefore a useful generic tool."

An RS teacher felt that value-added "was clearly better than league-tables" and an English teacher felt that it was "useful as a guide".

There were a range of concerns expressed by all those interviewed:

"Baseline data is too simplistic....it fails to account for differential development. It's parading as an exact science and it isn't" (an English teacher).

"The drawback at Key Stage 3 is that SATS do not tie up at with Key Stage 2. Level 4 at Key Stage 2 does not match level 4 at Key Stage 3. The levels need to be matched at each key stage" (a DT teacher).

"Value-added does not measure natural interest or curiosity...or whether a student has a questioning mind" (a Science teacher).

"It creates artificial figures and facts. It has some potential but many things cannot be measured numerically. The baseline data we use is very dicey and it is not very reliable in English" (an SEN teacher).

"It is hard to identify the effect of the school, let alone any individual teacher. Any baseline measure is bound to be simplistic.....it is probably not very suitable for aesthetic subjects" (a Mathematics teacher).

These concerns do raise questions about what value-added is measuring in reality, as opposed to what it claims to be measuring. However, despite these misgivings about value-added procedures in general, as the respondents saw CAT data as a more reliable baseline than KS2 data.

STAFF THOUGHTS AND FEELINGS ABOUT TARGETS

Attitudes towards targets, and the way that such information is used, are an important part of the consequential validity of these procedures. Since this kind of information is now also being used as part of the Performance Management apparatus, and since this also has implications in terms of moving up the post-threshold pay spine, it goes without saying that there are wide-ranging consequences which flow from the use of such procedures, for

staff and students. There was a wide variety of reactions to the idea of students being given targets:

"It's a good stick to beat them with...it makes assessment more automatic and easier" (a DT teacher).

"I didn't have targets when I was a student.....it would have made me very stressed. Now students seem to be hooked on levels and targets" (as RS teacher).

"We have so many distractions in teaching....targets help us to get back to what matters...it helps staff to focus" (a Mathematics teacher).

"Targets are worthwhile...it helps me to have an idea of what kinds might be capable of. I hear a lot of staff moaning about targets....more celebration is needed when targets are achieved" (an ICT teacher).

"It makes students more responsible for their own learning and gives staff something to hang their hats on" (a Science teacher).

"Staff are not happy about them especially if related to pay...there is cheating and manipulation of figures. It is very divisive" (an SEN teacher).

Concern was expressed by one teacher about the way in which some students used their targets "as a stick to beat teachers with". This is an interesting, and important observation, because what it suggests is that to some extent targets may be having precisely the opposite effect to that intended. The assumption underlying target-setting is that it should make students more responsible for their own learning and should motivate them to work harder. Clearly in a number of cases the pressure is being passed back to the teacher. As we have seen in the parental questionnaire, in the previous chapter, parents seemed remarkably relaxed about target-setting,

and this is strongly reinforced by the interviews with students, which we will see later in this chapter.

The fact that targets were never going to be very high for low-ability students was seen as quite a serious problem in terms of motivation and self-esteem but overall there was a strong sense that more able students liked targets, were motivated by them, and that targets promoted a healthy, competitive spirit.

Targets were seen by several respondents as "setting teachers up for blame" and "destroying a sense of team" and there was a feeling that departmental, or year-group targets would encourage teachers to work together, rather than against each other. Typical of the comments was:

"A good teacher will always strive to achieve the best results that they can for their students anyway.....there is no need for targets...a good teacher always sets their own targets" (a Modern Languages teacher).

It is interesting to set this alongside the views expressed by both students and parents in their questionnaires, in the previous chapter, where something like a quarter to a third of students and parents thought that targets were not necessary to motivate students to do their best. What this part of the discussion underlines is an important element in the use of target-setting procedures. If they are primarily used for measurement and judgement, rather than for building confidence, and celebrating success, overall attitudes towards them are likely to remain rather negative at best. Stress caused by teachers and students not meeting their targets was an underlying theme in many of the interviews.

With one exception, the view was that targets are generally good for the motivation of the middle and higher ability ranges, but a problem for those of lower ability. Competitive students were seen as "rising to the challenge" whilst less able students "think it is fairy dust and it will just appear". One teacher commented:

"Targets are implicit in middle-class family values...delayed gratification...so targets have very beneficial effects on the most able and little or no effect on the least able. The concept of working towards something over a long period of time is something they are not used to" (a Mathematics teacher).

Concern was expressed about the way that targets "set people up for failure" and about the way in which they "label children". Concern was expressed by virtually all respondents about the effects on lower ability children:

"They rarely have any hope of achieving their targets and this just demotivates them...it creates the very opposite effect to the one intended" (a Modern Languages teacher).

"Any motivation gains are outweighed by the increased anxiety and fear and even some of the more motivated are very anxious about targets" (an RS teacher).

It is interesting to contrast this with the views expressed in the student interviews which, as we will see later, suggests a rather more complex picture of student reactions. Also, as we saw in the previous chapter, the student and parental questionnaires suggest a remarkably relaxed attitude generally to targets. One teacher spoke of students seeing targets as their

due, not something they had to work for, echoing concerns noted earlier. A Science teacher, took a very different view:

"Targets are particularly useful for the least able because it causes them to make more effort, especially when techniques for bringing about improvement are identified in the classroom".

This last point is particularly telling because, as the analysis of the interviews with students reveals later, students seem to feel that they work harder in subjects where teachers give this kind of very specific advice about *how* to improve their work.

EFFECTS OF TARGETS ON STAFF MORALE, AND PERCEPTIONS OF THEMSELVES AND THEIR STUDENTS

Generally the views expressed were highly critical of the effects on teachers, although two respondents took a fairly positive line, and one adopted a neutral stance on the issue. Concerns were expressed about the amount of paperwork that targets, and tracking procedures, produce. A much more deep-rooted concern, however, was the way in which targets were being used to pressurize and judge staff. Typical of the comments were:

"The pressure is non-stop...I feel very cynical about targets" (an RS teacher).

"Targets are a management tool which arrived at a very bad time for the profession, just as it was moving into Performance Management, and performance-related pay. This has all been very divisive...it is seen as yet another burden" (an English teacher).

A DT teacher commented, in a very telling phrase:

"You don't make pigs fat by counting them…there is a real worry about the amount of paperwork it produces…it's just something else that has just been imposed on us".

An ICT teacher commented:

"Staff are working like mad to get the best results anyway….targets have increased this pressure even further but there often seems to be little commitment there in students".

This perception does seem to reflect the general ethos which emerged in the student and parent questionnaires in the previous chapter.

One teacher thought that staff gained satisfaction from achieving targets, especially with less able students, and another felt that targets "make staff review and challenge what they do". It is interesting to note that both of these comments were made by staff with management responsibilities. From a completely different perspective, the view was expressed that "if you can tell that some targets are inaccessible to students, it makes sense to treat them with caution". This seemed to echo a general feeling that staff treated targets for lower ability students with considerable caution, if not scepticism. The idea was put forward that perhaps student targets should be identified to staff, but not to students. This is an interesting idea in light of the concerns expressed by staff about lower ability students, students generally using targets to put pressure on staff, and the fact that about a quarter of students see themselves as working no harder as a result of having targets.

EFFECTS OF TARGETS ON TEACHING

Teaching to the test is an inevitable risk in a target-driven system of education. Since targets create further pressure within the system for summative types of assessment, the question arises, as to whether target-setting procedures are having a beneficial or deleterious effect on formative assessment in the classroom, on relationships in the classroom, and on the curriculum generally. The view was widespread that target-setting had distorted teaching in a considerable number of ways. Only one respondent offered much in the way of positive observations about the effects of targets on teaching, and interestingly this teacher, who has recently taken on more of a management role, indicated that very different comments would have been made about this issue twelve months earlier. Teaching to targets, tests, and assessment criteria, was seen as being a significant problem for virtually all teachers interviewed. Typical of the comments was:

"Learning is not as much fun as it used to be...creativity goes out of the window...I want to be able to teach – go off at tangents – let children relax and enjoy learning" (an SEN teacher).

"Education has become a commodity, a product, a means to an end as a result of having targets. Important stuff gets left out...kids are looking for short cuts to targets and are not inclined to think for themselves" (an ICT teacher).

A Mathematics teacher suggested:

"You feel vulnerable if you veer away from the syllabus...you don't take risks....there is a huge pressure to deliver a product".

Only one teacher felt that targets had not distorted teaching or classroom relationships. Other ideas were that disaffected children had become even

more disaffected as a result of having targets, that education had become much more mechanical, and that students "see qualifications as a way of promoting themselves". In trying to get an effective working relationship with students, one teacher felt that "targets can become an obstacle", particularly when testing, and retesting, to achieve targets, becomes the pattern in the classroom, which it often does at particular pressure points. The feeling that targets had really channelled pressure back onto teachers, rather than onto students, where it was rightly seen as belonging, was apparent in many of the interviews.

WIDER OBSERVATIONS AND SUGGESTIONS

A number of suggestions were put forward as to how we might do things differently and improve our professional practice. The idea of giving short-term targets to the less able was raised several times, as was the idea of giving more *credit* to students and staff for achieving their targets. There was concern that we should do something about the distorting effects of concentrating on borderline students because "we need to look good". The need for more diagnostic forms of assessment was identified, as was the need to be able to spend more time building better relationships with students, especially the less able. The idea of giving staff more support was mentioned a number of times, to avoid problems of staff feeling inadequate when they did not achieve their targets, especially when it was no fault of their own. From a broader perspective, one teacher (of English) felt that it was time that the teaching profession took a long look at itself and started challenging many of the philosophical issues underpinning value-added and target-setting procedures:

> "We need to assert our professional judgement and find leaders in the profession who are thinkers. There is a real need for a vision of what education should be about through dialogue with teachers....we need to remodel education so that it produces thinking people".

More appreciation of the work of teachers, especially at government level, was seen as being crucial, but sadly lacking. There was also a feeling amongst some teachers that the school needed to do much more to celebrate *achievement* in relation to targets. There should be "more cause for rejoicing" suggested one teacher, and another thought that, as a tutor, it would be much more productive to have time to praise and support students. On the other hand there was a feeling that the whole process was deskilling teachers and reducing autonomy and enjoyment. One teacher, of Mathematics, (with management responsibilities) observed:

"Target-setting has put assessment on a more rational and measurable kind of basis instead of just being wrapped up in words as it has been in the past. Standardisation makes comparisons much easier. This must be good for education".

Others clearly felt that this was the nub of the *problem*, rather than the nub of the *solution*, to improving education. Formative assessment was raised as a crucial issue in raising achievement, one teacher expressing the view that many staff seemed to have no real idea of what it is like for students trying to understand things, but that we can, if we try, "identify short-term things which can help to bring about immediate improvement". Once again, the feeling of imposition by government was raised, with the suggestion that the only way to cope with such managerialist attitudes was "pragmatic optimism". An SEN teacher commented:

"The whole target-setting process has left teachers paddling furiously and going nowhere. These processes should complement teaching, not get in the way".

This seems to sum up the core difficulties which many teachers feel that they face, working in a target-driven environment

COMMENTARY ON THE ANALYSIS OF STAFF INTERVIEWS

It would appear that most teachers appreciate the potential benefits of target-setting. However, it is also clear that most teachers find themselves concerned about the distorting effects on the curriculum, on their relationships with students, and on the relationships of students with them. The evidence is clear that teachers are concerned about the hugely bureaucratic machinery which target-setting, and tracking, gives rise to, and the fact that formative and diagnostic assessment, which is clearly at the heart of the business of raising performance, is being marginalised in the process. In general terms, support for targets is seen to increase with seniority. Furthermore, the conception of what these procedures might achieve, or indeed are achieving, becomes increasingly mechanistic with increasing seniority. The contrast with the attitude of teachers who spend most of their working lives in the classroom was stark at times. It is clear that for classroom teachers the fundamental concern is that they are dealing with human beings, some of whom can be extremely difficult, and that getting the best out of such a wide variety of students is far from being a mechanistic process. It calls on an immense repertoire of skills, knowledge, experience, intuition, optimism, and fortitude, not to mention the ability to cope with considerably increased pressures which come from senior managers, and parents, and from students themselves. The issue of how much additional pressure targets have put students under is interesting. As the details of the student interviews reveal later in this chapter, it would appear that there is substantial evidence to support the assertion made by a number of staff that teachers are carrying far more of the burden, and pressure, of student targets, than students are themselves.

INTERVIEWS WITH STUDENTS

GENERAL ATTITUDES TOWARDS TARGETS

The overall attitudes were very positive and there was a fairly consistent picture at the four different target levels, that students thought targets were a good thing, and that they made them work harder. Typical of the comments was:

"It makes me work harder – I'm disappointed when I don't get the level".

"Targets are good because...like...if your target is 6 and you get 5 it makes you work harder".

"Targets are good because if there's people that have problems they can get teachers to help them out...it will make their work go up".

One student, who was given a target level of 4, the lowest target that the school sets at Key Stage 3, indicated:

"It's sort of helped...I've been aiming to try to do better than the target".

This comment is clearly at odds with the observations of staff about the demoralising effects of low targets on low ability students, and about the lack of motivation on the part of such students. On the other hand, it could be argued, of course, that this comment cannot necessarily be taken as representative of all low ability students.

Whilst those with targets at level 7 were still supportive of the idea of having targets, they generally seemed to take a more dispassionate view of targets, perhaps because these students are usually highly motivated anyway. One student indicated that her level of entry, in one of her core

subjects, meant that she could not achieve her target in that subject. This is clearly a problem, given the different levels of entry in Mathematics and Science at Key Stage 3 if the school only sets one average target across all subjects. One student, with a target level of 6, complained about targets, observing:

"I don't reckon it's very good....you're expected to get the target...I don't like it....it puts you under too much pressure".

EFFECTS OF TARGETS ON MOTIVATION AND EFFORT

Almost consistently, at all target levels, students said that they worked harder in all three core subjects of English, Mathematics and Science, but that they did not really work harder in the non-core subjects, with the exception of the odd subject here and there. One student, at target level 4, said:

"I work harder to show that I can achieve higher than her target".

Another student commented:

"I really have to push myself hard in some subjects....because it's really what job you want to have".

This is an interesting concern for a student not yet fourteen years of age.

One student said that she felt that students with a target level of 5 believed it was easy to get, and that only students with targets at levels 6 or 7 really felt that they needed to work hard. This echoes the concerns of some of the staff that certain students seem to regard targets as what they *expect* to get, rather than what they had to *work for*. One student felt that they were simply not given enough work to do in non-core subjects, to enable them to meet their

targets, and this seemed to reflect an overall impression that whatever the pressures may be in nationally tested subjects, the atmosphere in teacher-assessed subjects at Key Stage 3 was much more pedestrian. Another common theme was the minimal relevance of the target in Years 7 and 8 by comparison with Year 9. Many students felt that there were no significant pressures on them until Year 9, either from teachers or from parents. Whilst most students felt quite strongly that targets had made them work harder, there was a general sense that targets produced a much higher level of motivation for those with a target level of 7, probably because of the potential prestige of taking extension papers and achieving level 8. This was summed up by a student at target level 7 who said:

"I work hard because I really want to achieve...I've worked harder than I did at primary school. It does actually help people to work harder...to get a higher level".

The question arises, of course, as to whether the high level of motivation for these students is the function of school targets, or of a much wider competitive attitude induced by their home background. Given that a quarter of students, in their questionnaires, said that they always worked as hard as they could, irrespective of whether they had targets or not, it is tempting to wonder whether the targets themselves fulfil the purposes other students, and presumably schools, think that they do.

ANXIETY ABOUT TARGETS

Surprisingly, given the staff concerns about anxieties and fears in students, in interviews, they almost universally said that they "never" worried about targets. One student at target level 4 made an interesting comment:

"If I had a really low target level, I would be worried, but I don't think mine is too low. It's sort of helped...I've been aiming to try to do better than the target".

Given that level 4 is the lowest target that the school sets at Key Stage 3, this is an interesting observation. Has the teacher skilfully protected her self-esteem by persuading her that there are lower targets? Has the lower level of the target produced a determination to "prove them wrong"? Whatever the reason, the motivational psychology of target-setting is clearly not straightforward. One student, at target level 5, said that she felt disappointed when first given the target in Year 7:

"It made me feel that teachers didn't expect much of me but I wanted to prove them wrong".

There seemed to be little or no increase in anxiety as a result of having targets, even in Year 9, where the pressures were obviously greater, at least in the core subjects.

CONVERSATIONS WITH PARENTS ABOUT TARGETS

Many of the students said that they had no idea what their parents thought about targets, because their parents never talked about the issue. One student said her parents simply encouraged her to do the best, and another said that her parents "sometimes ask questions – but they don't nag me". One student, at target level 4, said:

"My mum thinks they're a good thing...people...like it helps teachers to know what work to give you".

Another student said that her mother thought that targets were a good thing, but at other times thought that targets pressurized students too much.

"Sometimes I agree" the student opined. The overall impression, therefore, is of a fairly relaxed attitude by parents, confirming the findings of the parental questionnaire.

STUDENT CONVERSATIONS WITH EACH OTHER ABOUT PROGRESS TOWARDS THEIR TARGETS

The evidence was very clear that despite being given targets early in Year 7, after a short period of conversation about the targets, they are little talked about until Year 9. One student observed:

"When we first got them, in Year 7, we thought it's going to be in Year 9, so we didn't talk about it".

Given the pressures staff believe that the school is putting students under, this is somewhat surprising. It tends to bear out hints elsewhere in these interviews that perhaps students are not really being challenged very much in any subjects in Years 7 and 8. Generally, the only times that targets seem to be talked about is after tests, or internal examinations, or when reports are issued. The concept that an end-of-key-stage target is producing long-term motivation is clearly in question. For quite a number of students, the approach is fairly dispassionate:

"We mess around and say I'll get a better result than you...we just have a laugh about it".

When the question of whether students discuss their progress towards their targets with other students was pursued a little further, one student said:

"Yes...not...like to see who is best, but to see what other people have got. We like to see our results, to see who has got the highest".

Apparently, though, this did not lead to much in the way of competition between students. One target level 7 students said that students tended to compete against *themselves*, rather than against *each other.* The general tenor of these comments is therefore in rather stark contrast to the concerns and anxieties expressed in the staff interviews.

EXTENT TO WHICH TEACHERS HELP STUDENTS TO IMPROVE THEIR PERFORMANCE IN ORDER TO REACH THEIR TARGETS

Quite a number of students interviewed felt that they only got the kind of help which enabled them to raise their performance in about *half* of the subjects which they studied at Key Stage 3. Some said it was only a couple of subjects. A number of students said that they worked harder in the subjects where they got this kind of advice, and they also indicated that they wished that they were given this kind of help in *all* subjects. One student made quite a telling remark about the role of teacher, not just in giving advice, but in building confidence:

"Some targets, you know you will get there...others...you need a teacher to convince you".

One student said that "in some subjects we get no advice at all". In general, advice seemed to be better in the core subjects, with English and Mathematics gaining a particular mention by a number of students. Science was mentioned much less frequently.

EFFECTS OF HAVING TARGETS ON MOTIVATION ACROSS ALL SUBJECTS AT KEY STAGE 3

Not surprisingly, a number of students said that they worked harder in the three core subjects: "I work harder in those because they are more important". Non-core subjects were openly regarded as "less important" by

a number of students, whilst others said that they worked equally as hard in all of their subjects. One student said she definitely did not work as hard in subjects where she got little advice from teachers, whilst another said:

"I don't think I work as hard in Maths...I don't get on with the teacher".

PE was frequently identified as a subject that they regarded as not being very important, that they didn't like very much, and in which they were generally less determined, or concerned, about their target.

OVERALL ATTITUDES TOWARDS TARGETS

Of the eight students interviewed, seven thought that targets were a good thing. Supplementary questions to students seemed to indicate that this was not a trivial response, but that it had been well thought through. Typical further comments included:

"It does actually help people to work harder...they work harder to get a higher level".

"It makes you work harder...not just go at your own pace".

The general view was that other students, that they knew, held similar views "though I think that some get worried" as one student put it. One target level 5 student said:

"None of my friends lost heart when they got their target".

Another said that targets had helped her to surprise herself in terms of what she had achieved. The one student who took the opposing view observed:

"My target has not really made any difference to my attitudes, although it might be good to have a target at the end of Year7, Year 8 and Year 9".

This chimes in with earlier comments about the possibly higher effectiveness of shorter-term targets. How much does a target to be achieved at the age of fourteen mean to an eleven-year-old? The same student who was not enamoured of targets, paradoxically, also observed:

"In Year 9, you really have to try your best – do more hard work and stuff".

One student commented on the use of CAT data to set the targets in Year 7, commenting:

"How do you know...it's...like...pressure...sometimes you feel that you are not going to do your best in CATs... it's not really showing what we can do".

This echoes the concerns of a number of staff with regard to setting targets on the basis of a performance on one day at the beginning of Year 7. One student, at target level 6, said she thought that her target was a bit high at first, a theme that ran through a number of interviews, and the same student added:

"It upsets some people – it makes them fed up".

This suggests that the lack of anxiety and stress may be more apparent than real.

Students commented positively on the way that reports identified their progress towards their targets and it was clear that some students felt that if they were falling more than one level below their target, they needed to get

their skates on. "It makes you panic", was the observation of one student, whilst another added "you know where you need to try harder".

COMMENTARY ON THE ANALYSIS OF STUDENT INTERVIEWS

The overall support for targets amongst students was very surprising and student comments provided an interesting counterpoint to those of the staff. There was a strong impression that staff are correct in their perceptions that student targets have created far more pressures for teachers than they have for students. On the other hand, there was a persistent theme running through the student interviews that life was fairly relaxed in Years 7 and 8, and that whilst there were some pressures in core subjects in Year 9, on the whole there was not a great deal of pressure in non-core subjects. Some students even called for *more* work to be given to students in non-core subjects. It is always difficult, of course, to know just how pressured children might be feeling at this age, without necessarily being aware of it. Whilst there was a general sense that parents and teachers did not generally nag or pressurize, it was tempting to wonder whether more subtle emotional pressures might be at work, such as fear of falling below the target, or a sense that the school, or their parents, would be disappointed if they did not achieve their target. In the author's experience students often do not recognise this fear of disappointing somebody as a pressure, even at the age of seventeen or eighteen, until they begin to think about it, and yet fear of disappointing somebody can be a very powerful psychological pressure to work hard.

SUMMARY

The results of these interviews with students and staff, taken together with the analysis of the quantitative data, raise a number of wider questions with regard to the validity and reliability of value-added and target-setting procedures at Key Stage 3. There are some marked differences between the

perceptions of teachers, and those of students, with regard to targets. Taken together the evidence does tend to suggest that targets are creating far more pressure and stress for teachers than they are for students and real questions emerge about the extent to which these mechanisms increase motivation or achievement by students. As we saw in Chapter 3, the literature on targets has been written almost entirely from the perspective of government, educationalists, or teachers. Little research has been done with students themselves. This is a remarkable anomaly given that students are supposed to be the ultimate beneficiaries of targets.

In Chapter 6, therefore, we turn to the wider insights and implications which this study gives rise to.

CHAPTER 6

INSIGHTS AND IMPLICATIONS

It will be clear from the preceding two chapters that a considerable number of questions arise with regard to both the validity and reliability of value-added, and target-setting procedures. This, in turn, raises much wider questions with regard to consequential validity, such as accountability, the basis upon which policy is formulated, both at local and at national level, the purposes of assessment, and the philosophical assumptions which underpin the kinds of developments with which we are concerned. At the heart of all these wider issues is the question of the kinds of knowledge claims which are being made, and to what extent, if at all, these knowledge claims can be assumed to have any validity. If policy is to be formulated on any kind of rational basis we need to be very clear about the validity of the knowledge claims upon which it is founded.

THE VALIDITY AND RELIABILITY OF VALUE-ADDED PROCEDURES

We noted in Chapter 3 that for Messick (1993) validity is concerned with the degree to which empirical evidence and theoretical rationales support inferences and actions based on various assessment methods. *Judgement* and *inferences* are at the heart of this process and it is not an all-or-nothing concept. Inferences need to be tested, like hypotheses, both at a statistical and at a philosophical level. It is instructive, therefore, to look at the results of this research in terms of Messick's three categories of content, criterion-related, and construct validity.

Content validity is concerned with the issue of how representative assessment information is with regard to the domain it claims to be measuring. As we have seen in Chapter 3, there are some quite fundamental

content validity problems which arise even at individual subject level (Wiliam 1993) in trying to define a domain of knowledge. If we add to this the problems surrounding aggregation of scores across various areas of the domain (Wiliam 1995), we are left with a number of questions with regard to the content validity of assessment information at domain and subject level. Value-added procedures nearly always involve aggregating across several subjects (often the three core subjects) to produce a single figure which, theoretically, represents some kind of baseline measurement of the "ability" of any individual student. The very foundation, therefore, upon which value-added assumptions rest is itself extremely problematic. What, exactly, does a baseline figure actually represent? We have seen in this study how Key Stage 2 test data, as a baseline for value-added procedures, raises more questions than it answers, and even the kind of standardised measure which results from the CAT is not without its problems in terms of what it actually represents.

Criterion-related validity is concerned with the ways in which we use criteria to make inferences, based on assessment information, from one setting to another. This is a particularly critical validity issue for value-added because at the heart of value-added is the assumption that you *can* makes inferences from one situation to another with a reasonable degree of confidence and that you can even formulate assumptions about future performance in different settings. Yet we need to pause for only a moment to realise that students not only move from one school to another, and one teacher, or class-group, to another, but they are doing this at a time in their lives when their very awareness and perception of themselves and the world around them is continually changing, not least during their teenage years. If we add in to this equation the consequential effects of assessment, value-added, and target-setting procedures, which we will examine later, we clearly face a state of affairs in which the validity of generalising from one setting to another, for individual students, is seriously open to question. As Wiliam (1996) asks, how can we separate out performance from context? If

this is a problem, as Wiliam suggests, in terms of setting standards, because it inevitably involves a degree of arbitrariness in terms of resolving conflicting priorities, how much more of a problem is it in terms of value-added procedures? We have seen in this study how the extent to which baseline data can explain the variation in subsequent performance data is often quite limited, and that this often leaves the question wide open as to what other factors may be accounting for the variation in subsequent performance. We also saw in Chapter 3 the problems explored by Goldstein (2001) in terms of how different value-added models produce widely different outcomes in terms of judgements about performance. In effect, like models of national economic performance, what answers you get out of the model depend critically upon what information you put in, and what you put in depends on all sorts of assumptions about what factors you believe to be important or critical. It is, perhaps, this area of value-added, above all, which needs a great deal more research and explication. Government and schools, as well as teachers and parents, need to understand, at a much more fundamental level, the assumptions upon which these procedures are founded, and the very real limitations and dangers which are inherent in the outcomes themselves, and even more fundamentally in the actions which are taken on the basis of such outcomes. The knowledge claims which are made on the basis of value-added procedures are problematic because the procedures which give rise to such "knowledge" are problematic.

Messick's third area of concern is construct validity. As we saw with intelligence, if you cannot define, with any degree of certainty, what intelligence is, it is extremely difficult to measure it. Construct validity is a highly contested notion in value-added and we noted in Chapter 3 how the concept is very slippery (Saunders 1999). Does it represent value added by the student or the institution? Is it a measure of progress or prediction? Are there simple linear relationships between different sets of performance data for individuals, classes, schools, or even nationally? The concept seems to change its shape and form, like a blob of mercury in a dish, depending on

how it is being used, and from what perspective one is viewing it. From the point of view of validity theory in educational assessment, if construct validity is problematic, then everything based on such a construct will be too. As we have seen in this research, the extent of this problem may well vary from one subject area to another. It seems reasonably clear from the local and national data which we examined in Chapter 4 that for Mathematics, value-added has quite a high level of construct validity because of its scalable nature in terms of its domains (Wiliam 1993). Because competence at lower levels is a prerequisite of achieving competence at higher levels, it is possible to formulate a reasonably effective definition of what progress, and therefore value-added, might represent in Mathematics. However, this is much more problematic in Science, even more so in English, and for the other National Curriculum subjects, as we have seen, there would appear to be little or no rational basis upon which to formulate a definition as to what value-added actually represents in these areas.

The consequential effects of assessment discussed by Messick are often little understood, or even ignored, but they do have a critical bearing on validity, as we saw in Chapter 3. Any kind of assessment information not only provides feedback-loops to teachers and learners, with important consequences in terms of self-perception and confidence, but it also, over a period of time, establishes the rules of the game which both teachers and students learn to exploit to their advantage. Because value-added information supposedly represents quite fundamental information about the performance of individual students, teachers, and schools, based on national comparisons, the feedback effects of this information are likely to be considerable, not just because it purports to represent judgements about *past* performance, but also because it is used as the basis for making predictions, or setting targets, for *future* performance. As we will see in the discussion of target-setting, this can have important consequential effects on future performance, particularly for the most able, and the least able, students.

Little attention seems to have been paid in the literature to the issue of the consequential effects of value-added procedures though the issues raised by Saunders (2000) with regard to the ways in which value-added data is used in schools, as we have seen, would seem to suggest that consequential effects cannot, and should not, be ignored.

What this suggests, overall, is that there are serious validity problems where value-added is concerned. If we look at the eight links in the validity chain suggested by Crooks et al (1996) we see that the validity of value-added procedures are open to question on the basis all eight criteria, but particularly with regard to aggregation, generalisation, extrapolation, evaluation, decision-making and impact. The claims made on the basis of value-added are so fundamental that they inevitably call into question the very basis of the assessment procedures upon which they are based, let alone the inferences which are made from this kind of data. Since Crooks et al believe that weakness in any one of these eight links threatens the validity of any procedure overall, it is clear that with the validity of all eight of these links being open to question in value-added procedures, the validity of value-added overall must be in serious doubt.

Because value-added procedures involve making generalisations over long time periods (typically two or three years or more), reliability is as critical an issue in value-added as validity issues are. As we have seen in both the local and national data, there are considerable problems with regard to the reliability of both test data, and teacher assessment data. Considerable fluctuations are apparent year on year, in many subjects, and this, as much as the validity problems, perhaps represents the most serious question facing those who advocate the use of value-added procedures. If either the baseline data, or the subsequent performance data, cannot be shown to be reasonably consistent over time, then we have little or no basis upon which we can rationally make any judgements about progress, or compare classes or schools. As we have seen, Feldt and Brennan (1993) have identified a

number of significant factors which impinge on reliability in a *stable* system of assessment. In an *unstable* system, a whole range of additional factors come into play which raise the most fundamental questions about the basis upon which we can make any judgements of value-added. Tymms and FitzGibbon (2001) have raised a number of questions about the reliability of national test data, suggesting that "the changes seen between 1995 and 1999 are so dramatic and so out of step with other longitudinal data as to raise serious questions about their being true representations of changes in standards" (p.162). They report that no anchor tests are used to check standards across the years "as one might expect" (p.162), and that, perhaps of even more concern, data on the possible consequences of decisions made at meetings to decide on level and grade boundaries in national tests, are made available to those taking those final decisions. At best this makes decisions about final boundaries less than completely objective. At worst this at least opens up the possibility of a charge of manipulation of results to serve the needs of national targets. Tymms and FitzGibbon argue, with some justification, that without some kind of standardised national tests, against which National Curriculum test standards can be measured year on year, reliability must be in doubt. "In order to monitor standards we need a benchmark that does not alter" (p.168). They argue that evidence from standardised entrance examinations used by some university departments lends support to the idea that there is some kind of grade inflation occurring year on year. Furthermore, the evidence they have accumulated over several years from schools suggests that students are working less hard, and are being given less homework, than they have in the past. They raise particular concerns about standards in Key Stage 2 tests and this has been highlighted further in the study conducted by CEM (2002). In a comparison of results on a standardised test of ability (PIPS) for 5000 students a year between 1997 and 2002, with National Curriculum test results at Key Stage 2, there appears to be a serious mismatch between the claims being made nationally for improvements in performance in English at Key Stage 2, and performance on the standardised test. Whilst there is some question with

regard to what aspects of English PIPS and Key Stage 2 are actually testing, it raises a question, at the very least, as to how reliable National Curriculum test results can be assumed to be. Interestingly, as we saw earlier, improvements in Mathematics at Key Stage 2 do seem to be supported by the PIPS data reported in this study.

These concerns are lent further support by the investigations of McCarty et al (2001) who compared National Curriculum test levels of individual students with results obtained from on-line testing of some 4000 students in primary and secondary schools in 2001. Their findings suggest that the differences in performance "force pertinent questions about the compatibility of standards between subjects, as well as between the key stages" (p.12) and that "continual changes make it exceedingly difficult to provide a common measuring scale that is robust, when there is no stable baseline to build upon" (p.13).

THE VALIDITY AND RELIABILITY OF TARGET-SETTING PROCEDURES

If the validity and reliability of value-added procedures are in doubt, it follows by implication that target-setting procedures are too, because targets are virtually always formulated on the basis of value-added analysis, and predictions of future performance, as well as on local and national comparisons. It hardly needs stating that if the reliability of the data used in value-added procedures is in question, any targets based upon such data must also be of questionable reliability. In terms of the validity of such targets perhaps the two most critical questions are how suitable any given target is for a particular student, and secondly what the consequential effects of such targets are for students and teachers.

The evidence from this study is rather mixed. It appears that the policy of setting one, high-challenge, target average, across all subjects, does play at

least some positive role in raising standards. Students, parents and teachers generally seem to regard this as a useful tool, and a reasonable compromise. On the other hand, the fact that some 10% of students often do not know what their target is, even after three years study at Key Stage 3, invalidates the use of such procedures for these students. It also emerged that some 25% of students believe that they work as hard as they can anyway, and therefore feel that they do not need targets. As we noted in the staff interviews, staff effectively ignore targets for less able students very often, because of the potentially demotivating effects of trying to persuade students to work hard to achieve what is, in reality, a very low target. Whilst some students at Key Stage 3 said that being given a low target at the beginning of the Key Stage provided a spur to prove the school wrong, it is tempting to wonder to what extent this is a motivational device instilled in students by teachers, rather than any intrinsic part of the student's own motivation. The evidence from staff was certainly stronger for Key Stage 4, where students being given a High Challenge Target Average of grade E at the beginning of the Key Stage were quite demoralised, in a number of cases.

A wider question arises with regard to the validity of targets in education and that is the question of the effects on the teaching and learning process itself. The assumptions which underpin the supposed value of target-setting are essentially both behaviourist and reductionist, yet there seems to be little or no research which explores the validity of these assumptions. The assumption seems to be that if you set a target, students and teachers will automatically do everything within their power to achieve it. Is this assumption justified? To some extent it may be, but one of the consequential effects is that we see a narrowing down of the concept of education to a concern with targets and results, and a failure to recognise the wider effects of this change in perceptions and attitudes (Flecknoe 2001). We saw in the staff interviews how many students are solely concerned with achieving targets by the quickest and most efficient route, and that knowledge is

increasingly becoming a commodity in the eyes of students, rather than something of intrinsic value, echoing Lyotard's (1984) postmodernist view of knowledge. Skimming over the surface, and getting to the destination as quickly as possible are increasingly overtaking any pleasure in the journey itself, and staff saw the whole process as tending to produce an increasingly mechanistic approach in the classroom. On the other hand staff, students, and parents did seem to regard targets as a useful benchmark by which to make some judgements about what level of achievement might be possible.

Evidence from this study seems to lend support to the view of Fielding (1999) that targets distort the teaching and learning process. Staff reported less risk taking, less exploration of the highways and byways of subject areas, and an increasing pattern of teaching to the test, rote-learning and teaching of *habits* rather than *understanding*. Furthermore the evidence which emerged that target-setting had produced considerably increased pressures for teachers, but relatively little for students, raises some quite basic questions about the validity of target-setting procedures. Targets are supposed to raise the performance of *students* as well as *teachers*. If students are merely seeing their target as something they have a *right to expect* the school, and individual teachers, to deliver, as the evidence from this study suggests, then the value of giving targets to students must be in doubt.

Wider evidence is beginning to emerge of both the dangers and the potential shortcomings of target-setting procedures. We have seen a number of cases in the last two or three years of schools, and even individual headteachers, interfering with examination papers, and other forms of data, in a fraudulent attempt to meet performance targets, either local or national. It is reasonable to assume that, at a less serious level, schools will at least attempt to massage or manipulate their data, if they can do this within the rules, as Fielding (1999) suggests. The introduction in 2002 of so-called "booster classes" by the government, in an open attempt to move borderline students

from Level 4 to Level 5, not only represents a further attempt to manipulate statistics to achieve published targets, but it sits very uneasily alongside government claims to be committed to social inclusion when such programmes are only funded for those students who will improve national and local performance figures.

One question with regard to target-setting which appears to be so obvious that nobody is asking it is whether we believe that there is a finite limit with regard to the extent to which we can raise standards. Is there is a ceiling nationally? Do individuals have their own ceiling above which they cannot rise? Do teachers have some kind of ceiling of professional competence above which they cannot rise? We do not have empirical answers to these questions, and perhaps we never will, but experience tends to suggest that there are finite limits for all of us in terms of what we can achieve, no matter how much we may want to achieve something, and no matter how hard we may try. The evidence from the NFER chances tables would appear to suggest that at lower ability levels, at least, some students have reached a plateau beyond which they are unlikely to go. The results of the so-called "progress tests", where students who only achieve Level 3 in their Key Stage 2 tests, resit them early in Key Stage 3, lend further support to this possibility, and it raises the question as to whether we are going to see this plateau effect spreading upwards through the ability range as targets become ever more challenging year on year. This is not just a question of ability. It is also the case that the extent to which we achieve any particular goal is dependent upon what other, competing, goals we may be trying to achieve at the same time. Teenage years are a time when there are many competing and conflicting goals, and it must be open to question as to whether merely setting targets for students substantially changes levels of motivation or effort. The evidence from the student interviews for this study is not particularly encouraging.

Evidence is beginning to emerge of increasing disenchantment with targets as a result of the year on year pressures to raise standards. The TES (2002) reported that the government was being left in no doubt that "targets were among heads' main complaints" (p.20) and that "many headteachers – and a few education authorities – are openly rejecting these diktats from above" (p.20). This has partly been triggered by the failure to reach some national targets in 2001 and the potential shortfall of 2% - 3% on the targets for 2002. The view of the TES is that headteachers are therefore professionally justified in treating targets as "aims not edicts. They know they cannot perform miracles" (p.20) especially given the teacher shortages. It is tempting to wonder, given how ambitious government targets are for 2007, whether the entire target-setting policy will be abandoned, particularly if the targets look increasingly unobtainable. It would certainly be a brave and unusual political party which pressed on with such a policy if the problems which are beginning to emerge become more evident, despite the TES (2002) reports that the Secretary of State for Education would "rather aim too high and miss than set expectations too low" (p.20). It seems fairly clear that what is emerging from this debate are fundamental questions about the kind of education system that we want, raised by Tikly et al (2000). The effects of globalisation, global competition, the shockwaves going through the major capitalist institutions as a result of recent revelations, and the general questioning of the economic system in which we find ourselves, all raise quite profound questions about the nature of the education system which we have created since Callaghan's speech in 1976. It is entirely possible that disillusionment with a target-driven education system, increasingly designed to harness economic competition with the rest of the world, will gather pace. If it does, we may see a major restructuring of educational philosophy in the next two decades.

PERFORMATIVITY, ASSESSMENT, AND ACCOUNTABILITY

We noted in Chapter 2 how Lyotard (1984) sees knowledge as having become nothing more than a commodity in our post-modern culture, and furthermore how knowledge is fundamentally based in the kinds of language games which we play, involving a series of moves and counter-moves as in a game of chess. Performativity is one of these language games, according to Lyotard, in which the need to establish some kind of "truth" is replaced by the need to find "the best possible input-output equation" (p.46). With the collapse of grand-narratives this new type of "knowledge" is used to augment power because with the power of new technology "performativity increases the ability to produce proof" (p.46). Given the findings of this study, and the wider literature on assessment and value-added, it is difficult not to view what is increasingly emerging in educational assessment as a clear example of the kind of performative knowledge which Lyotard has identified. National tests and value-added calculations seem fundamentally more concerned with "proving" that certain things are happening in the education system, than with diagnosing what is actually going on, and the power of new technology is increasingly being harnessed, as Lyotard suggested, to augment these procedures. We have already seen how these processes are tending to distort the teaching and learning process, and the attitudes and conceptions of teachers and students with regard to their role and purpose within the education system, producing the kind of mercantilized knowledge described by Lyotard. Education, according to Lyotard, is now legitimized by performativity, and the need to optimize skills "designed to tackle world competition" (p.48), and to provide the social system with players "capable of acceptable fulfilling their roles" (p.48). Performativity increasingly tends to subordinate educational institutions to existing powers in society, rather than emancipating the learner and creating ideals.

It is clear that the ever-increasing tendency for assessment to mean little more than sets of numerical data, by which we measure the performance of the "system", is fundamentally changing the very nature of what goes on in schools and classrooms. We have seen in Chapter 3 the increasing concern in the last decade in terms of the extent to which all forms of assessment have become dominated by summative, rather than formative types of assessment, and value-added procedures, and target-setting procedures in particular, take this process a stage further. Foucault (1979), was of the view that examinations embody "the techniques of an observing hierarchy and those of a normalizing judgement" (p.184), but if we take closer look at his concept of what he called panopticism, we see that the kind of surveillance and normalizing judgements about which he was so concerned have increased dramatically with the kinds of procedures we have been examining in this study. His idea arises from Bentham's architectural design in which a central, circular tower has windows opening onto cells in which "a madman, a patient, a condemned man, a worker or a schoolboy" (p.200) can be kept under constant surveillance. The purpose of this, according to Foucault is to "induce in the inmate a state of conscious and permanent visibility that assures the automatic functioning of power" (p.201). If we look at the ways in which we are now using technology, even at school and national level, let alone at the individual level, to track individual performance, and predict the likelihood of meeting predetermined targets, we can see that we have devised a form of educational panopticism which is terrifying in terms of its power to control. Knowledge and power have become even more inextricably linked than perhaps even Foucault (1980) could have imagined. Many schools now, including the one which has been the subject of this study, routinely track the performance of individual student on a week-by-week basis, in much the same way as many industries use technology to provide themselves with an up-to-the-minute picture of where they are in terms of input and output, and in terms of costs and profits.

Broadfoot (2001) expresses serious concern about this discourse of performativity, suggesting that the rationalistic assumption that it is possible and desirable to measure the performance of individuals and organisations seems to have largely gone unchallenged in education. Broadfoot sees the "rampant growth of a forest of assessment procedures" (p.137) which "threatens to throttle the whole education system" (p.137). Assessment is now almost entirely conceptualised as a measuring device, "the results of which are used to goad" (p.143). Performativity has replaced empowerment in assessment. For Broadfoot this ethos in education, driven by the threat of international competition, produces "an increasing pre-occupation with managerialism" (p.151) and as educational institutions come to accept this discourse, "the domination of the discourse of 'performativity' is daily reinforced" (p.151). It is difficult, therefore not to see targets in education as a clear example of Baudrillard's simulacrum because they increasingly seem to bear little or no relation to the reality which they are supposed to represent, and have taken on a life of their own which masks the very absence of this relationship with any reality.

We have already examined, in Chapter 3, the concerns with regard to the accountability problems which this ethos produces, but accountability has generally been understood in terms of the performance of schools and individual teachers. What we see emerging from all this is increasing accountability pressures on students and parents as well, who now have access to "information" which encourages them to be making detailed, and on-going evaluations of their performance. It is an open question as to what the long-term effects of this kind of surveillance will be on children, or on their attitudes towards education. As we saw earlier, there are signs of cracks in the educational superstructure with regard to targets. Will we see the same attitudes emerging in parents and students?

GENERALISABILITY AND TRANSFERABILITY

Schofield (1993), in examining the issue of the possible generalisability of any piece of research, suggests three areas of concern. To what extent is the study typical, common, or ordinary? To what extent does the site of the study fit with future likely trends and issues? To what extent is the site of the study exceptional or ideal? The contextual information given about the site of this study, in Chapter 1, indicates very clearly how, on a number of factors, the site of this study is entirely typical. There are two respects in which it is not necessarily typical however. One is that it is a single-sex girls school, and secondly it is a school which is showing significant improvements at Key Stage 3, rather than the "dip" which many schools are experiencing at this level. As a state comprehensive school which has a good record of implementing new developments, it is a site which is likely to show a close fit with future trends and issues. To what extent the site of the study is exceptional, or ideal, is rather more difficult to judge. Clearly the school has been performing very well in terms of benchmark comparisons, but as we have seen, the changing nature of the way the school is categorised in benchmark terms is beginning to change perceptions about the extent to which the school is performing as well as might have been supposed.

Guba and Lincoln's (1989) criteria of credibility, transferability, dependability and confirmability are perhaps even more useful in terms of estimating the extent to which these findings might be applicable to other institutions, or in terms of raising questions about national data and procedures. In terms of credibility, this study has fulfilled many of the criteria suggested by Guba and Lincoln with respect to faithful reconstructions, particularly prolonged engagement with the site. The burden of proof in terms of transferability is on the receiver according to Guba and Lincoln "and depends entirely on the degree to which salient conditions overlap or match" (p.241). The database of information provided

in Chapter 1 is designed, as Guba and Lincoln recommend, to facilitate transferability judgements by others. Dependability is concerned with tracking changes over time, and documenting these changes. Given that this research represents a five-year longitudinal study of value-added, and a three-year study of target-setting, it would seem reasonable to claim a high level of dependability in this study. Confirmability is concerned with the extent to which data are rooted in contexts, and are not merely figments of the researcher's mind. Given the commitment of this study to open up insights, and Lyotard's intersections of narratives, however conflicting these may be, and given also the biographical context of the author given in Chapter 1, and the raw data in the appendices, it is perhaps for the reader to judge to what extent the confirmability criteria have been met by this study.

CONCLUSIONS

One of the aims of this study was to open up insights into the issues of value-added and target setting procedures at Key Stage 3 by examining a number of different avenues of investigation. As we have noted, since this is a case-study based on one particular site, it is for others to judge to what extent the findings are transferable and therefore no specific recommendations have been made. In Chapter 2 we noted Scott's (2000) commitment to a form of transcendental realism, rejecting the concept that knowledge is entirely socially constructed in favour of certain emergent ontological properties being identifiable. What is clear from this study is that value-added is a very problematic as a concept, let alone as a procedure. Given that target-setting procedures are built upon such a problematic base, they too provide us with more questions than answers. It seems fairly clear that both of these procedures are changing the nature of education, not just at school and national level, but also at the level of the individual student and teacher. Some of these effects might appear to be positive, others appear to be more detrimental. At a wider philosophical level, such procedures are beginning to prompt deep-rooted questions about the nature of the education

system we have created in response to the call of Prime Minister Callaghan in 1976 for a Great Debate, and the national angst about education seems to be as endemic now as it was twenty six years ago.

APPENDIX 1

INITIAL QUESTIONNAIRE ON TARGETS FOR YEAR 7

Name_____ Tutor-group_____

1 Write down what you think your High Challenge Target Average Level is:

Level_____

2 How did you feel when you were first given your target? (Circle ONE answer):

Very anxious Anxious A little bit anxious Not anxious at all

3 How well did you understand what having a target meant when it was first given to you? (Circle ONE answer):

I didn't understand at all I think I understood I clearly understood

4 How did you feel about trying to meet your target when it was first given to you? (Circle ONE answer):

The target was too low The target was about right The target was too high

5 Have you talked to your tutor about your target? (Yes or No):_____

6 Have you talked to your parents about your target (Yes or No):_____

7 How likely do you think it is that you will reach your target? (Circle ONE answer):

I don't think I will reach my target in any of my subjects

I think I might reach my target in a few of my subjects

I think I might reach my target in most of my subjects

I think I might reach my target in all of my subjects

8 Overall, what do you think about the idea of having targets? (Circle ONE answer):

I don't think targets will make <u>any difference</u> to how I work in school

I think targets <u>might</u> encourage me to work harder in some subjects

I think targets will encourage me to <u>work harder in all of my subjects</u>

APPENDIX 2

QUESTIONNAIRE ON TARGETS FOR END OF YEAR 9

Name_____ Tutor-group_____

1 Write down what you think your High Challenge Target Average Level has been for Key Stage 3: Level_____

2 How do you feel now about the target which you were given? (Circle ONE answer):

The target was too low The target was about right The target was too high

3 How do you feel now about having one target for all of your subjects? (Circle ONE answer):

The target was not right for any of my subjects

The target was only right for a few of my subjects

The target was about right for most of my subjects

The target was right for all of my subjects

4 To what extent did having a target affect the way that you worked? (Circle ONE answer):

I didn't work harder in <u>any subject</u> as a result of having a target

I worked harder in <u>some subjects</u> as a result of having a target

I worked harder in <u>all of my subjects</u> as a result of having a target

5 How often have you talked to your <u>tutor</u> about trying to reach your target? (Circle ONE answer):

Never Once or twice Reasonably frequently Very frequently

6 How often have you talked to your <u>subject teachers</u> about trying to reach your target? (Circle ONE answer):

Never Once or twice Reasonably frequently Very frequently

7 Overall, how do you feel about targets? (Circle ONE answer):

No use Sometimes helpful Very helpful

APPENDIX 3

STAFF QUESTIONNAIRE ON TARGETS

We have been using High Challenge Targets for two years now. In order to establish where this is working well, and where we need to make changes or improvements, please tick the relevant box for each of the factors on the questionnaire.

	Poor	Fair	Good	Very good	Excellent	Cannot decide
Mechanism for deciding targets (CATs, chances tables, ALIS)						
How realistic are the targets for your subject?						
How well do students understand the "high challenge" aspect of targets?						
Extent to which targets help to improve learning						
Extent to which targets help to improve teaching						
Attitudes towards targets by students						
Attitudes towards targets by parents						

Please add below any further comments which you wish to make:

APPENDIX 4

QUESTIONNAIRE ON TARGETS TO PARENTS

Please circle ONE answer to each of the following questions.

1 How familiar would you say you were with your daughter's High Challenge Target?

Not familiar at all

I think I know what the target is

I am quite clear about what the target is

2 How successful do you think the school has been in explaining how your daughter's target was calculated?

Very poor Poor Satisfactory Generally good Very good

3 Who do you feel about the High Challenge Target which was given to your daughter?

The target was too low

The target was about right

The target was too high

4 How do you feel about the idea of your daughter having one target for most subjects?

The target is not right for any subjects

The target is only right for a few subjects

The target is about right for most subjects

The target is appropriate for all subjects

Please turn over

5 To what extent do you think having a target has affected the way that your daughter works?

I don't think my daughter worked harder in any subject as a result of having a target

I think my daughter worked harder in some subjects as a result of having a target

I think my daughter worked harder in all subjects as a result of having a target

I don't think my daughter needs a target – she always does her best anyway

6 How often would you say that you have talked to your daughter about progress towards her targets in the last academic year?

Never Once or twice Reasonably frequently Very frequently

7 Overall, how do you feel about the idea of your daughter having targets?

Targets are not helpful

Targets are sometimes helpful

Targets are always helpful

8 How do you feel about the reports you receive from the school with regard to your daughter's progress towards her High Challenge Target?

Reports are not very helpful

Reports are reasonably clear

Reports are very informative

Please add below any further comments which you would like to make:

APPENDIX 5

QUESTION PROMPTS FOR INTERVIEWS WITH STUDENTS

These questions were used as prompts for semi-structured interviews with Year 9 students. Supplementary questions were formulated as a result of the responses given to these questions.

1 What do you think about the idea of having targets?

2 Do you think targets have made you work harder?

3 Do you worry about having targets?

4 What do your parents say to you about targets?

5 Do you compare your targets with those of other students?

6 Do teachers tell you how to improve your work in order to achieve your targets?

7 Do you work equally as hard in all of your subjects in order to achieve your targets?

8 Do students talk to each other about progress towards their targets?

9 Overall, do you think targets are a good thing?

10 Are there any other thoughts or observations you would like to make?

APPENDIX 6

QUESTION PROMPTS FOR INTERVIEWS WITH STAFF

These questions were used as prompts for semi-structured interviews with staff. Supplementary questions were formulated as a result of the responses given to these questions.

1 Do you think value-added is a coherent concept?

2 What are your thoughts about targets for students?

3 What are your thoughts about targets for staff?

4 What effects, if any, do you think targets have had on staff?

5 What effects, if any, do you think targets have had on teaching?

6 Is there anything that we should be doing differently?

7 Are there any other thoughts or observations that you would like to make?

APPENDIX 7

1994 INTAKE

RESULTS OF CAT

(LEVEL F, TAKEN IN 1996)

SS = Standard Age Score (Mean = 100)

	Verbal SS	Quantitative SS	Non-verbal SS	Average SS
94/1	91	84	90	88
94/2	120	100	98	106
94/3	108	89	95	97
94/4	83	76	73	77
94/5	130	119	124	125
94/6	88	87	90	88
94/7	94	106	106	102
94/8	83	99	108	97
94/9	94	85	105	95
94/10	103	98	103	101
94/11	130	128	127	129
94/12	92	81	89	87
94/13	80	73	94	82
94/14	114	97	105	105
94/15	100	103	117	107
94/16	93	92	108	98
94/17	99	95	98	97
94/18	125	100	110	112
94/19	84	88	101	91
94/20	94	92	95	94
94/21	113	95	101	103
94/22	114	130	125	123
94/23	109	107	113	110
94/24	92	80	94	89
94/25	89	105	109	101
94/26	92	92	103	96
94/27	106	108	115	110
94/28	113	102	108	108
94/29	92	102	99	98
94/30	130	107	113	117
94/31	109	116	123	116
94/32	130	97	100	109
94/33	110	104	118	111
94/34	94	106	94	98
94/35	109	107	113	110
94/36	111	104	115	110
94/37	120	106	121	116
94/38	120	104	110	111
94/39	102	87	93	94
94/40	78	75	88	80
94/41	93	96	96	95
94/42	96	115	111	107

94/43	92	109	108	103
94/44	91	97	105	98
94/45	97	97	108	101
94/46	112	95	119	109
94/47	93	102	113	103
94/48	96	76	72	81
94/49	84	98	94	92
94/50	84	90	106	93
94/51	79	85	108	91
94/52	96	88	95	93
94/53	70	74	75	73
94/54	83	73	87	81
94/55	99	98	103	100
94/56	122	89	116	109
94/57	70	71	81	74
94/58	104	80	103	96
94/59	104	93	100	99
94/60	89	79	87	85
94/61	128	127	119	125
94/62	130	118	119	123
94/63	99	77	94	90
94/64	102	105	111	106
94/65	129	128	121	126
94/66	105	91	101	99
94/67	120	91	115	109
94/68	100	88	99	96
94/69	93	81	88	87
94/70	99	83	97	93
94/71	101	100	114	105
94/72	92	93	103	96
94/73	102	108	99	103
94/74	76	77	99	84
94/75	103	98	107	103
94/76	85	92	98	92
94/77	97	97	79	91
94/78	85	82	96	88
94/79	116	112	115	114
94/80	105	89	91	95
94/81	78	83	80	80
94/82	121	118	129	123
94/83	109	107	110	109
94/84	95	104	86	95
94/85	117	115	117	116
94/86	108	95	103	102
94/87	117	122	120	120
94/88	96	113	111	107
94/89	96	99	96	97
94/90	88	80	85	84
94/91	99	108	125	111
94/92	111	105	125	114
94/93	117	106	115	113
94/94	115	127	117	120
94/95	107	96	107	103
94/96	104	89	92	95
94/97	87	99	98	95

94/98	108	104	127	113
94/99	70	80	94	81
94/100	105	94	100	100
94/101	108	98	99	102
94/102	120	120	124	121
94/103	117	114	117	116
94/104	98	93	122	104
94/105	85	75	83	81
94/106	102	78	96	92
94/107	81	78	88	82
94/108	82	77	95	85
94/109	125	115	113	118
94/110	112	89	110	104
94/111	103	89	99	97
94/112	98	116	110	108
94/113	96	92	106	98
94/114	98	104	116	106
94/115	81	70	78	76
94/116	94	105	101	100
94/117	89	102	94	95
94/118	97	100	113	103
94/119	88	85	101	91
94/120	106	93	103	101
94/121	95	86	109	97
94/122	89	100	106	98
94/123	116	94	115	108
94/124	98	96	94	96
94/125	122	106	113	114
94/126	90	95	87	91
94/127	87	95	97	93
94/128	122	117	118	119
94/129	108	94	108	103
94/130	114	110	110	111
94/131	72	76	92	80

APPENDIX 8

1994 INTAKE

KEY STAGE 3 RESULTS: CORE SUBJECTS (1997)

TEST AND TEACHER ASSESSMENT LEVELS

TA = Teacher assessment Eng = English Ma = Mathematics Sci = Science

	Eng Test	Eng TA	Ma Test	Ma TA	Sci Test	Sci TA	Average Test
94/1	4	4	4	4		4	4.00
94/2	7	7	6	7	6	6	6.33
94/3	7	6	6	6	6	6	6.33
94/4	4	2	4	4	4	4	4.00
94/5	8	8	7	7	7	7	7.33
94/6	4	4	5	4	4	4	4.33
94/7	5	6	6	5	5	5	5.33
94/8	4	5	5	5	5	5	4.67
94/9	5	4	5	5	6	6	5.33
94/10	6	5	6	6	5	5	5.67
94/11	7	7	8	8	7	7	7.33
94/12	4	4	4	4	5	5	4.33
94/13	4	4	4	4	4	4	4.00
94/14	6	6	5	6	5	5	5.33
94/15	5	6	7	7	6	6	6.00
94/16	4	5	6	6	5	5	5.00
94/17	4	5	5	4	5	5	4.67
94/18	7	7	6	6	7	7	6.67
94/19	4	3	5	4		4	4.50
94/20	4	5	4	5	5	5	4.33
94/21	5	5	5	5	5	5	5.00
94/22	6	6	8	8	7	7	7.00
94/23	5	6	6	6	6	7	5.67
94/24	4	3	4	4	4	4	4.00
94/25	4	5	6	6	5	5	5.00
94/26	4	4	6	6	5	5	5.00
94/27	5	5	6	7	5	5	5.33
94/28	5	6	6	6	6	6	5.67
94/29	4	4	6	6	5	7	5.00
94/30	6	7	7	7	7	7	6.67
94/31	5	6	7	7	6	6	6.00
94/32	8	8	5	5	6	6	6.33
94/33	5	5	6	6	5	5	5.33
94/34	5	5	6	6	5	5	5.33
94/35	4	6	7	7	6	6	5.67
94/36	6	6	7	7	6	6	6.33
94/37	5	6	7	7	5	6	5.67
94/38	6	7	6	7	6	6	6.00
94/39	6	5	6	6	5	5	5.67
94/40	5	4	4	3	4	4	4.33
94/41	4	4	5	5	5	5	4.67
94/42	5	5	7	7	6	6	6.00

94/43	5	5	5	5	5	5	5.00
94/44	4	4	4	4	5	5	4.33
94/45	4	5	5	6	5	5	5.00
94/46	5	6	6	7	6	6	5.67
94/47	5	5	6	6	5	6	5.33
94/48	5	6	4	4	5	5	4.67
94/49		4		5	4	4	4.00
94/50	4	4	5	5		4	4.50
94/51		4		4	4	4	4.00
94/52	4	5	4	4	4	4	4.00
94/53	3	2	2	3	3	3	2.67
94/54	4	4	4	4	5	5	4.33
94/55	4	5		5	5	5	4.50
94/56	5	7	6	6		6	5.50
94/57		3		2	3	3	3.00
94/58	5	5	4	4	5	5	4.67
94/59	5	5	5	5	6	6	5.33
94/60	4	4	3	3	5	5	4.00
94/61	6	7	8	8	7	7	7.00
94/62	6	7	7	8	7	7	6.67
94/63	4	4	4	4	4	4	4.00
94/64	5	6	6	6	6	6	5.67
94/65	6	7	8	8	7	7	7.00
94/66		4		5		5	
94/67	6	6	6	6	7	7	6.33
94/68	4	4	5	5	5	5	4.67
94/69	4	3	3	3	4	4	3.67
94/70	4	5	5	5	5	5	4.67
94/71	6	6	7	7	5	6	6.00
94/72	5	4	4	4	5	5	4.67
94/73	5	5	5	5	5	5	5.00
94/74	4	3	4	4	4	4	4.00
94/75	5	6	6	6	5	5	5.33
94/76	4	4	5	6	5	5	4.67
94/77	4	5	5	5	4	4	4.33
94/78	4	4	4	4	5	4	4.33
94/79	7	7	7	7	6	6	6.67
94/80	5	5	5	5	6	6	5.33
94/81	4	2	3	3	5	5	4.00
94/82	6	6	7	8	6	6	6.33
94/83	5	7	6	6	5	5	5.33
94/84	5	5	5	5	5	5	5.00
94/85	5	6	7	7	6	6	6.00
94/86	5	5	6	6	6	6	5.67
94/87	5	7	7	8	7	7	6.33
94/88	4	5	6	7	6	6	5.33
94/89	5	5	6	6	6	6	5.67
94/90	4	3	4	4	4	4	4.00
94/91	4	5	6	6	5	5	5.00
94/92	5	6	6	6	7	7	6.00
94/93	7	7	7	7	6	6	6.67
94/94	6	7	8	8	7	7	7.00
94/95	5	6		5	6	6	5.50
94/96	5	6	4	4	5	5	4.67
94/97	5	5	5	5	5	5	5.00

94/98	6	6	6	6	5	6	5.67
94/99	4	3	4	4	3	3	3.67
94/100	6	6	6	6	5	5	5.67
94/101	5	5	5	5	5	5	5.00
94/102	7	8	7	7	6	6	6.67
94/103	6	6	7	7	6	6	6.33
94/104	6	6	6	6	5	6	5.67
94/105	4	4	4	4	4	4	4.00
94/106	5	5	5	5	5	5	5.00
94/107	4	3	4	4	5	5	4.33
94/108	4	3	4	4	5	5	4.33
94/109	7	7	7	7	7	7	7.00
94/110	5	5	6	6	5	5	5.33
94/111	4	5	6	6	6	6	5.33
94/112	5	5	6	6	5	5	5.33
94/113	4	5	6	6	5	5	5.00
94/114	5	5	5	5	5	5	5.00
94/115	4	2	3	3	3	3	3.33
94/116	5	6		5		4	5.00
94/117	5	5	5	5	4	4	4.67
94/118	5	5	6	6	6	6	5.67
94/119	5	4	4	4	4	4	4.33
94/120	6	6	6	6	5	5	5.67
94/121	4	5		5	4	5	4.00
94/122	4	5	6	6	3	5	4.33
94/123	6	6	6	6	5	5	5.67
94/124	5	5	5	5	5	5	5.00
94/125	6	7	7	7	6	6	6.33
94/126	5	4	5	5	5	5	5.00
94/127		4	4	5	4	4	4.00
94/128	8	8	7	7	6	6	7.00
94/129	6	6	6	6		5	6.00
94/130	5	5	6	6	5	5	5.33
94/131	4	2	3	3	4	4	3.67

APPENDIX 9

1994 INTAKE

KEY STAGE 3 RESULTS: NON-CORE SUBJECTS (1997)

TEACHER ASSESSMENT LEVELS

DT = Design Technology Geo = Geography Hi = History

ICT = Information and Communications Technology

MFL = Modern Foreign Languages Mu = Music PE = Physical Education

	DT	Geo	Hi	ICT	MFL	Art	Mu	PE
94/1		5	4	5	3	B	B	A
94/2	6	7	5	5	5	A	C	B
94/3	6	7	6	5	5	A	C	B
94/4	4	4	4	5	3	C	A	B
94/5	6	7	7	5	6	A	D	B
94/6	3	5	4	5	3	C	B	B
94/7	5	5	4	5	4	A	C	B
94/8	3	5	4	5	4	B	A	B
94/9	5	6	5	5	4	B	B	B
94/10	5	5	6	5	5	B	A	B
94/11	7	7	7	5	6	A	D	B
94/12	5	5	5	4	4	B	A	B
94/13	5	5	3	4	4	B	A	B
94/14	5	6	5	5	5	B	B	B
94/15	5	6	6	5	5	A	C	B
94/16	5	6	4	5	4	B	B	B
94/17	2	4	5	5	4	C	A	B
94/18	7	7	6	5	6	A	C	B
94/19	2	4	3	3	4	B	A	B
94/20	5	5	4	5	4	B	B	B
94/21	5	6	5	5	4	A	A	B
94/22	7	6	6	5	5	B	C	B
94/23	6	7	5	5	5	B	B	B
94/24	4	4	4	5	4	B	B	B
94/25	5	4	5	5	4	A	A	B
94/26	5	5	5	5	5	C	A	B
94/27	7	7	4	5	5	B	B	B
94/28	6	6	5	5	5	B	B	B
94/29	4	5	3	5	6	A	B	B
94/30	5	6	7	5	5	A	C	B
94/31	5	7	5	5	5	A	B	B
94/32	6	7	7	5	5	A	B	A
94/33	6	6	5	5	4	B	B	B
94/34	6	5	5	5	4	B	A	B
94/35	7	7	5	5	5	A	D	B
94/36	7	7	6	5	5	A	D	B
94/37	7	7	6	5	5	A	B	B
94/38	7	7	7	5	6	C	D	B

94/39	5	6	6	5	5	A	B	B
94/40	6	4	4	4	4	B	B	B
94/41	6	5	4	5	5	A	C	B
94/42	7	7	4	5	6	C	B	B
94/43	5	5	4	5	4	B	B	B
94/44	3	4	4	5	4	B	B	B
94/45	5	5	5	5	4	B	B	B
94/46	6	5	5	5	5	A	B	B
94/47	6	6	5	5	4	B	B	B
94/48	5	5	5	5	5	B	A	B
94/49	3	4	3	5	3	C	A	A
94/50	4	5	4	5	4	B	B	B
94/51	6	4	4	4	3	A	A	B
94/52	3	5	4	4	4	B	A	B
94/53	3	4	3	4	3	C	B	B
94/54	5	5	4	5	4	C	A	B
94/55	4	4	4	5	4	A	A	B
94/56	5	6	5	5	5	A	A	B
94/57		4	3	3	3	B	A	A
94/58	6	4	4	5	4	B	B	B
94/59	5	6	7	5	5	A	B	B
94/60	3	4	4	4	4	B	B	B
94/61	5	7	6	5	5	B	C	B
94/62	7	7	7	5	6	A	D	B
94/63	5	5	4	5	4	A	B	B
94/64	4	5	5	5	5	A	C	B
94/65	6	7	6	5	6	C	B	B
94/66	5	6	4	5	4	A	A	B
94/67	5	6	5	5	5	A	B	B
94/68	5	5	5	5	5	B	B	B
94/69	3	4	3	4	3	A	A	A
94/70	5	7	6	5	5	B	C	B
94/71	7	7	5	5	6	A	C	B
94/72	6	5	5	5	4	B	A	B
94/73	5	5	4	5	5	B	B	B
94/74	1	4	4	5	4	B	A	A
94/75	4	5	4	5	5	A	B	B
94/76	6	5	4	5	4	A	C	B
94/77	4	6	4	5	4	B	C	B
94/78	4	4	4	4	4	B	A	B
94/79	5	7	7	5	6	A	D	B
94/80	5	6	4	5	4	B	B	A
94/81	1	4	4	3	3	C	A	A
94/82	7	7	6	5	6	C	D	B
94/83	4	5	5	5	4	C	B	B
94/84	5	5	5	5	5	A	B	B
94/85	5	6	6	5	5	A	C	B
94/86	5	7	4	5	5	A	A	B
94/87	6	7	6	5	6	A	D	B
94/88	5	6	4	5	5	B	B	B
94/89	4	6	4	5	3	B	B	B
94/90	4	5	4	4	4	B	B	B
94/91	6	6	5	5	4	B	C	B
94/92	6	6	6	5	5	A	B	B
94/93	6	6	6	5	6	A	D	B

94/94	6	7	6	5	6	A	C	B
94/95	5	5	4	5	4	B	A	B
94/96	1	7	5	5	5	B	B	B
94/97	4	4	4	5	4	B	B	B
94/98	6	7	5	5	5	A	B	B
94/99	3	4	3	3	3	B	A	A
94/100	6	7	6	5	6	B	C	B
94/101	5	5	5	5	5	A	C	B
94/102	5	6	6	5	5	A	A	B
94/103	5	6	5	5	5	A	C	B
94/104	7	6	6	5	5	B	D	B
94/105	5	4	4	4	4	B	B	B
94/106	5	5	5	5	5	B	B	B
94/107	4	4	4	5	3	B	A	A
94/108	4	5	4	5	4	B	B	B
94/109	6	7	7	5	6	B	D	B
94/110	4	6	4	5	3	C	B	B
94/111	4	6	5	5	4	B	C	B
94/112	7	6	4	5	5	B	B	B
94/113	6	5	5	5	5	B	B	B
94/114	7	6	4	5	4	B	B	B
94/115	5	5	4	5	4	B	A	B
94/116	5	5	4	5	4	B	B	B
94/117	6	5	5	5	5	C	B	B
94/118	6	5	5	5	4	A	B	B
94/119	4	5	4	5	4	C	A	B
94/120	5	6	4	5	5	C	C	B
94/121	6	5	5	5	5	A	A	B
94/122	6	5	4	5	5	B	C	B
94/123	6	7	6	5	5	A	D	B
94/124	3	5	4	5	5	C	B	B
94/125	6	7	7	5	5	A	D	B
94/126	7	6	4	5	5	B	B	B
94/127	4	5	5	5	4	A	A	B
94/128	7	7	7	5	6	B	C	B
94/129	6	6	6	5	5	B	A	B
94/130	7	6	4	5	5	A	C	B
94/131	5	4	3	3	3	B	A	B

APPENDIX 10

1995 INTAKE

RESULTS OF CAT

(LEVEL E, TAKEN IN 1996)

SS = Standard Age Score (Mean = 100)

	Verbal SS	Quantitative SS	Non-verbal SS	Average SS
95/1	98	88	103	96
95/2	100	92	108	100
95/3	98	95	88	94
95/4	97	89	99	95
95/5	129	112	105	115
95/6	96	84	104	95
95/7	96	91	100	96
95/8				
95/9	79	86	82	82
95/10	89	92	107	96
95/11	120	110	127	119
95/12	87	91	97	92
95/13	100	95	115	103
95/14	100	100	94	98
95/15	105	98	110	104
95/16	107	100	115	107
95/17	86	83	95	88
95/18	104	107	111	107
95/19	109	95	107	104
95/20	89	91	97	92
95/21	104	88	101	98
95/22	123	108	119	117
95/23	109	105	117	110
95/24	97	100	109	102
95/25	111	115	108	111
95/26	103	108	110	107
95/27	92	87	113	97
95/28	98	98	110	102
95/29	114	97	103	105
95/30	110	110	112	111
95/31				
95/32	126	120	111	119
95/33	116	104	108	109
95/34	81	85	89	85
95/35				
95/36	112	112	112	112
95/37	115	103	111	110
95/38	105	106	106	106
95/39	102	75	97	91
95/40	120	104	106	110
95/41	115	87	105	102
95/42	106	108	105	106

95/43	98	109	116	108
95/44	113	95	102	103
95/45	90	92	97	93
95/46	107	*105	112	108
95/47	109	108	111	109
95/48	98	96	108	101
95/49	76	78	70	74
95/50	91	95	106	97
95/51	117	130	128	125
95/52	111	95	110	105
95/53	105	87	102	98
95/54	121	104	110	112
95/55	84	76	83	81
95/56	110	99	105	105
95/57	106	108	130	115
95/58	91	81	93	88
95/59	77	85	77	80
95/60	99	88	91	93
95/61	104	94	97	98
95/62	103	91	105	100
95/63	119	94	108	107
95/64	117	122	128	122
95/65	119	94	108	107
95/66	93	84	84	87
95/67	116	85	92	98
95/68	85	85	82	84
95/69	97	96	106	100
95/70	107	114	119	113
95/71	115	85	105	102
95/72	83	76	89	83
95/73	104	85	79	89
95/74	115	123	109	116
95/75	117	110	107	111
95/76	102	92	94	96
95/77	99	94	108	100
95/78	76	80	78	78
95/79	129	112	119	120
95/80	112	101	110	108
95/81	90	99	106	98
95/82	98	101	101	100
95/83	82	96	106	95
95/84	85	91	106	94
95/85	84	80	82	82
95/86	74	81	71	75
95/87	82	93	106	94
95/88	117	98	111	109
95/89	85	103	97	95
95/90	108	104	108	107
95/91	105	111	107	108
95/92	95	100	105	100
95/93	86	85	99	90
95/94	85	70	81	79
95/95	111	104	119	111
95/96	96	116	120	111
95/97	88	88	105	94

95/98	117	117	119	118
95/99	88	89	95	91
95/100	87	86	100	91
95/101	122	113	130	122
95/102	84	88	99	90
95/103	84	80	97	87
95/104	92	76	95	88
95/105	113	92	106	104
95/106	102	89	88	93
95/107	82	86	97	88
95/108	108	110	103	107
95/109	120	109	108	112
95/110	122	118	113	118
95/111	86	89	82	86
95/112	117	111	117	115
95/113	79	75	83	79
95/114	103	92	107	101
95/115	78	78	85	80
95/116	117	109	110	112
95/117	126	92	113	110
95/118	83	86	86	85
95/119	74	86	71	77
95/120				
95/121	88	84	99	90
95/122	120	113	120	118
95/123	92	103	113	103
95/124	93	84	95	91
95/125	98	102	103	101
95/126	95	98	103	99
95/127	127	117	106	117
95/128	102	94	111	102
95/129	87	82	101	90
95/130	86	105	100	97
95/131	77	80	85	81
95/132	117	115	111	114
95/133	80	78	93	84
95/134	91	86	102	93
95/135	110	92	100	101
95/136	101	104	97	101
95/137	115	108	108	110
95/138	108	109	108	108
95/139	97	82	95	91
95/140	123	115	117	118
95/141	100	94	113	102
95/142	94	99	94	96
95/143	85	99	104	96
95/144	108	113	107	109
95/145	89	101	107	99
95/146	118	113	124	118
95/147	93	89	92	91
95/148	81	73	94	83
95/149	109	92	115	105
95/150	90	80	96	89
95/151	103	94	95	97
95/152	92	108	107	102

95/153	108	115	128	117

APPENDIX 11

1995 INTAKE

KEY STAGE 2 RESULTS: CORE SUBJECTS (1995)

TEACHER AND TEST ASSESSMENT LEVELS

TA = Teacher assessment En = English Ma = Mathematics Sci = Science

	Eng Test	Eng TA	Ma Test	Ma TA	Sci Test	Sci TA	Average Test
95/1	4	4	4	4	4	4	4.00
95/2	4	4	3	3	4	4	3.66
95/3	3	3		3	3	4	3.00
95/4							
95/5							
95/6	3	3	3	2	4	2	3.33
95/7	3	4	3	3	4	4	3.33
95/8	4	4	3	3	4	3	3.66
95/9		3		3		2	
95/10							
95/11	4	5	5	5	5	5	4.66
95/12							
95/13	4	4	3	3	4	4	3.66
95/14	4	5	4	4	4	4	4.00
95/15	4	4	3	4	4	3	3.66
95/16	4	4	4	4	5	4	4.33
95/17	3	4	3	3	4	3	3.33
95/18	3	4	4	4	5	4	4.00
95/19							
95/20	4	4	4	4	5	4	4.33
95/21	4	4	3	4	4	4	3.66
95/22	5	5	4	4	4	4	4.33
95/23	4	5	4	5	4	4	4.00
95/24	3	3	4	4	5	5	4.00
95/25	5	5	5	5	5	5	5.00
95/26	4	4	4	4	4	4	4.00
95/27	3	4	3	4	5	4	3.66
95/28	4	3	3	3	4	4	3.66
95/29	4	4	3	4	5	5	4.00
95/30	4	4	4	4	4	4	4.00
95/31		3	3	3	3	3	3.00
95/32	4	5	4	5	4	4	4.00
95/33	4	4	4	3	4	3	4.00
95/34	3	3	3	3		3	3.00
95/35	4	5	4	4	5	4	4.33
95/36	4	4	4	4	4	4	4.00
95/37	4	5	4	4	4	4	4.00
95/38	5	5	4	4	5	4	4.66
95/39							
95/40	4	4	5	4	5	4	4.66
95/41	5	5	4	4	5	5	4.66
95/42	4	4	4	4	5	4	4.33

95/43	4	4	4	4	4	4	4.00
95/44	4	4	3	3	4	4	3.66
95/45	3	3	3	3	4	3	3.33
95/46	5	5	4	4	4	4	4.33
95/47	4	5	4	4	4	4	4.00
95/48							
95/49	2	2	2	2	2	2	2.00
95/50	4	4	5	5	5	5	4.66
95/51	5	5	4	4	4	4	4.33
95/52							
95/53	4	4	3	3	4	3	3.66
95/54							
95/55	3	3		3	3	3	3.00
95/56	4	4	3	3	4	4	3.66
95/57	4	4	4	4	4	4	4.00
95/58	3	4	3	4	3	4	3.00
95/59		3	3	3	4	4	3.50
95/60	4	3	3	3	4	3	3.66
95/61	4	4	4	4	4	4	4.00
95/62	4	4	4	4	4	4	4.00
95/63	5	5	4	4	5	5	4.66
95/64	4	4	4	4	5	4	4.33
95/65	4	4	3	3	4	3	3.66
95/66	3	3		3		3	3.00
95/67	4	4	4	3	4	4	4.00
95/68		3		3		4	
95/69	3	4	4	4	5	4	4.00
95/70	5	5	5	4	5	4	5.00
95/71	4	4	3	3	4	4	3.66
95/72	3	3	3	3	3	3	3.00
95/73							
95/74	4	5	5	4	4	5	4.33
95/75	5	5	5	5	5	4	5.00
95/76	5	5	4	4	5	4	4.66
95/77	3	4	3	2	3	2	3.00
95/78							
95/79							
95/80	5	5	5	5	4	4	4.66
95/81		3		3		3	
95/82	4	4	4	3	3	3	3.66
95/83	3	4	3	4	3	3	3.00
95/84	2	3	3	3	3	3	2.66
95/85	3	3	2	2	3	3	2.66
95/86	2	2	3	3	3	3	2.66
95/87	3	3	3	3	3	3	3.00
95/88	4	5	4	4	5	5	4.33
95/89	3	3	3	3	4	3	3.33
95/90	3	4	4	4	4	4	3.66
95/91	4	5	3	4	4	4	3.66
95/92	4	4	4	4	4	4	4.00
95/93	3	3	3	3	4	3	3.33
95/94	3	3	3	4	4	4	3.33
95/95							
95/96	4	4	5	5		4	4.50
95/97	3	3	3	3	3	3	3.00

95/98	4	4	5	5	4	4	4.33
95/99	4	4	3	3	5	4	4.00
95/100							
95/101	4	4	5	5	5	4	4.66
95/102	3	3	3	3	4	3	3.33
95/103	3	2		2	3	2	3.00
95/104	3	4	3	4	4	4	3.33
95/105	4	4	4	4	4	4	4.00
95/106							
95/107	3	3		3	3	4	3.00
95/108	4	4	4	3		4	4.00
95/109	4	3	4	4	4	3	4.00
95/110							
95/111	4	4	3	3	4	4	3.66
95/112	4	4	4	4	4	4	4.00
95/113	3	3	2	2	3	3	2.66
95/114	4	4	4	4	4	4	4.00
95/115		2		1		2	
95/116	4	5	5	5	5	5	4.66
95/117	5	5	4	4	5	5	4.66
95/118	3	3		3	3	2	3.00
95/119							
95/120	3	3	2	3	3	3	2.66
95/121	4	4	3	3	4	4	3.66
95/122	4	4	4	4	4	4	4.00
95/123	4	4	4	3	4	3	4.00
95/124	4	3	3	3	4	3	3.66
95/125	3	3	3	3	4	3	3.33
95/126	4	3	4	3	4	3	4.00
95/127	5	6	5	5	5	5	5.00
95/128	3	4	4	4	4	5	3.66
95/129	3	3	3	3	4	4	3.33
95/130	3	3	4	4	5	4	4.00
95/131	3	3		3	3	3	3.00
95/132	4	5	4	5	4	5	4.00
95/133	3	3	3	2	3	2	3.00
95/134	3	3	3	3	4	4	3.33
95/135	4	4	3	4	4	4	3.66
95/136	4	4	3	4	4	4	3.66
95/137	5	5	4	4	5	5	4.66
95/138	4	4		4		3	4.00
95/139	4	4	3	3	4	3	3.66
95/140							
95/141							
95/142	3	3	3	3	5	4	3.66
95/143	2	2	3	2	3	2	2.66
95/144	3	4	4	4	4		3.66
95/145	3	3	3	3	4	3	3.33
95/146							
95/147	3	3	3	3	4	3	3.33
95/148	3	4	3	3		4	3.00
95/149		4		3		4	
95/150	3	4	3	3	4	4	3.33
95/151	4	3	4	4	5	4	4.33
95/152	4	4	4	4	4	4	4.00

95/153	4	4	5	4	4	3	4.33

APPENDIX 12

1995 INTAKE

KEY STAGE 3 RESULTS: CORE SUBJECTS (1998)

TA = Teacher assessment Eng = English Ma = Mathematics Sci = Science

	Eng TA	Eng Test	Ma TA	Ma Test	Sci TA	Sci Test	Average Test
95/1	5	5	5	6	5	5	5.33
95/2	6	6	5		5		6.00
95/3	4		3	3	3		3.00
95/4	4	4	5	4	4		4.00
95/5	7	8	7	7	7	7	7.33
95/6	5	5	5	5	5	6	5.33
95/7	4	5	4	4	5	5	4.66
95/8	5	5	5	5	5	5	5.00
95/9	4	4	4	4	4	3	3.66
95/10	3	3	5		4	4	3.50
95/11	7		7	8	7	7	7.50
95/12	3		4		3	3	3.00
95/13	5	6	6	6	6	5	5.66
95/14	6	7	5	6	5	5	6.00
95/15	3	3	5	5	5	4	4.00
95/16	6	6	6	6	6	6	6.00
95/17	4	4	4	4	6	5	4.33
95/18	6	6	6	6	5	5	5.66
95/19	4		4	5	5	4	4.50
95/20	5	5	5	5	5	4	4.66
95/21	5	6	5	5	5	6	5.66
95/22	6	5	7	7	7	6	6.00
95/23	6	6	7	7	7	6	6.33
95/24	5	5	6	7	4	5	5.66
95/25	5	5	6	6	6	6	5.66
95/26	5	4	6	6	5	5	5.00
95/27	5	5	6	6	5	5	5.33
95/28	5	6	5	5	5	5	5.33
95/29	6	7	5	5	6	5	5.66
95/30	7	7	6	6	5	6	6.33
95/31	4	4	3	4	4	4	4.00
95/32	7		7		7		
95/33	6	6	6	6	6	6	6.00
95/34	5	5	4	4	4	4	4.33
95/35	6	6	5	5	5	6	5.66
95/36	6	6	6	6	5	5	5.66
95/37	6	6	7	6	6	6	6.00
95/38	7	6	7	7	6	7	6.66
95/39	6	6	4	5	5	5	5.33
95/40	6	6	7	7	6	6	6.33
95/41	5	5	5	5	6	6	5.33
95/42	5	5	6	6	5	6	5.66
95/43	5	5	7	6	5	6	5.66
95/44	5	5	5	4	5	5	4.66

95/45	5	5	5	5	4	4	4.66
95/46							
95/47	6	6	6	6	7	6	6.00
95/48	5	6	6	7	6	6	6.33
95/49	3		2		3		
95/50	4	4	6	6	5	5	5.00
95/51	6	6	7	7	6	7	6.66
95/52	5	5	6	6	5	6	5.66
95/53	6	6	5	4	5	5	5.00
95/54	7	7	6	6	7	5	6.00
95/55	4	4	3	3	4	4	3.66
95/56	5	5	6	6	5	5	5.33
95/57	6	6	7	7	6	6	6.33
95/58	4	5	4	4	4	4	4.33
95/59	4	4	4	4	4	4	4.00
95/60	6	6	4	4	5	5	5.00
95/61	6	6	6	6	5	6	6.00
95/62	7	6	6	6	5	6	6.00
95/63		6	6	6	6	6	6.00
95/64	7	6	7	7	7	7	6.66
95/65	5	6	5	5	5	4	5.00
95/66	4	4	4		4	3	3.50
95/67	5	6	5	5	5	6	5.66
95/68	3		3	3	4	4	3.50
95/69	5	6	5	5	5	6	5.66
95/70	7	6	7	7	7	6	6.33
95/71	5	5	5	6	6	6	5.66
95/72	4	4	4	4	4	3	3.66
95/73	5	7	5	4	5	5	5.33
95/74	7	6	6	7	5	7	6.66
95/75	5	6	6	7	7	6	6.33
95/76	5	6	4	5	5	5	5.33
95/77	3		4	4	5		4.00
95/78	3		4		4		
95/79	7	8	7	7	7	7	7.33
95/80	7	7	7	7	6	6	6.66
95/81	4	4	5	5	4	4	4.33
95/82	5	6	6	5	5	5	5.33
95/83	5	5	5	5	5	5	5.00
95/84	4	4	4	4	4	4	4.00
95/85	5	5	2	3	3	3	3.66
95/86	2		3		3	3	3.00
95/87	3	4	4	4	4	3	3.66
95/88	8	7	8	7	7	7	7.00
95/89	5	5	5	6	5	5	5.33
95/90	5		6	6	5	5	5.50
95/91	5	6	6	6	5	5	5.66
95/92	5	4	5	5	4	5	4.66
95/93	5	5	3	4	4	4	4.33
95/94	4	4	3	3	5	4	3.66
95/95	6	6	6	6	7	6	6.00
95/96	7	5	8	8	7	7	6.66
95/97	4	5	4	4	4	5	4.66
95/98	6	6	7	7	7	7	6.66
95/99	5	5	4	4	5	4	4.33

95/100	4	4	5	5	5	5	4.66
95/101	4	5	7	7	6	6	6.00
95/102	4	4	4	4	4	4	4.00
95/103	3		4	4	4	4	4.00
95/104	4		4	4	4	4	4.00
95/105	6	7	6	6	5	5	6.00
95/106	5	4	5	5	5	5	4.66
95/107	5	5	4	4	4	4	4.33
95/108	6	6	6	6	7	6	6.00
95/109	6	6	6	7	7	6	6.33
95/110	8	8	7	7	7	6	7.00
95/111	5	5	4	4	4	4	4.33
95/112	7	7	6	6	7	5	6.00
95/113	5	4	3	3	4	4	3.66
95/114	5	5	5	6	5	5	5.33
95/115	4		3	3	4	3	3.00
95/116	6	6	8	8	7	7	7.00
95/117	6	6	6	6	6	6	6.00
95/118	4	4	4	4	4	3	3.66
95/119	3		2	3	3	2	2.50
95/120	4	5	3	3	5	5	4.33
95/121	4	5	5	4	4	4	4.33
95/122	6	6	7	7	7	6	6.33
95/123	5	5	7	7	5	6	6.00
95/124	4	5	5	5	5	5	5.00
95/125	5	4	5	5	5	5	4.66
95/126	5	5	6	6	5	6	5.66
95/127	7	6	8	8	7	7	7.00
95/128	6	6	5	5	5	5	5.33
95/129	5	4	3	3	4	4	3.66
95/130	5	5	5	5	5	5	5.00
95/131	4	3	4	4	4	4	3.66
95/132	8	6	7		5		6.00
95/133	4	4	4	4	4	4	4.00
95/134	5	6	5	5	5	5	5.33
95/135	5	6	5	5	6	6	5.66
95/136	7	7	6	6	5	6	6.33
95/137	6	6	6	6	6	6	6.00
95/138	6	6	6	6	5	5	5.66
95/139	5	6	4	5	5	5	5.33
95/140	7	8	7	7	7	7	7.33
95/141	6	7	6	6	6	5	6.00
95/142	5	5	5	5	5	5	5.00
95/143	4		5	5	4	4	4.50
95/144	6	6	7	7	6	6	6.33
95/145	5	5	4	5	4	5	5.00
95/146	6	6	8	7	7	7	6.66
95/147	5		4		4		
95/148	4		4	4	4	4	4.00
95/149	4	5	5	5	5	5	5.00
95/150	4	5	3	4	4	4	4.33
95/151	4	5	5	5	5	5	5.00
95/152	6	6	6	6	5	5	5.66
95/153	6	6	7	7	7	6	6.33

APPENDIX 13

1995 INTAKE

KEY STAGE 3 RESULTS: NON-CORE SUBJECTS (1998)

TEACHER ASSESSMENT LEVELS

DT = Design Technology Hi = History Geo = Geography Mu = Music

ICT = Information and Communications Technology

MFL = Modern Foreign Languages PE = Physical Education

	DT	ICT	Hi	Geo	MFL	Art	Mu	PE
95/1	3	5	5	4	3	C	B	
95/2	7	5	7	5	5	C	B	
95/3	3	4	4	3	2	B	A	
95/4	5	5	5	4	3	C	B	
95/5	7	6	7	7	6	B	C	
95/6	7	5	6	5	5	C	B	
95/7	5	5	5	6	3	C	C	
95/8	4	5	5	5	5	C	A	
95/9	6	3	4	3	2	B	A	
95/10	3	4	4	4	2	B	A	
95/11	7	7	7	7	6	D	D	
95/12	3	4	5	4	4	B	B	
95/13	7	5	6	5	5	B	C	
95/14	5	5	6	6	3	B	C	
95/15	3	5	4	4	3	C	A	
95/16	7	5	6	5	5	B	B	
95/17	3	3	4	4	3	B	B	
95/18	6	5	6	6	5	B	B	
95/19	4	5	5	4	3	B	A	
95/20	6	5	5	5	4	C	A	
95/21	6	5	6	6	5	C	B	
95/22	5	6	7	6	5	C	C	
95/23	7	6	6	6	5	B	C	
95/24	7	5	6	5	3	B	A	
95/25	6	5	6	5	5	B	B	
95/26	5	5	5	4	3	B	B	
95/27	5	4	6	7	4	B	C	
95/28	5	5	5	6	4	C	B	
95/29	7	5	7	7	6	C	C	
95/30	7	5	7	6	5	C	B	
95/31	5	5	6	5	5	C	B	
95/32	7	7	7	6	6	B	D	
95/33	6	6	6	7	5	C	B	
95/34	5	4	5	4	3	B	B	
95/35	5	5	5	6	4	A	A	
95/36	6	6	6	6	5	B	B	
95/37	7	5	7	7	6	C	D	
95/38	7	7	7	7	6	B	D	

95/39	3	5	4	5	3	B	A	
95/40	5	5	6	6	5	C	C	
95/41	5	4	5	6	4	B	B	
95/42	7	6	7	6	5	C	B	
95/43	6	7	6	5	5	C	C	
95/44	5	5	5	4	3	B	B	
95/45	5	4	5	4	4	B	B	
95/46	6	7	7	6	6	C	D	
95/47	6	7	7	7	5	C	C	
95/48	6	5	5	6	4	B	B	
95/49	6	5	6	5	5	C	C	
95/50	3	5	5	4	3	B	C	
95/51	5	7	6	6	5	C	C	
95/52	6	6	6	6	4	C	B	
95/53	6	6	6	6	5	B	B	
95/54	7	5	7	7	5	C	C	
95/55	5	3	5	4	2	B	A	
95/56	5	5	4	4	4	B	B	
95/57	7	7	7	6	5	C	C	
95/58	6	4	5	4	4	B	B	
95/59	4	4	6	4	2	B	B	
95/60	3	5	4	4	2	B	A	
95/61	7	5	7	7	5	C	C	
95/62	6	5	7	6	6	B	C	
95/63	3	4	4	3	3	C	A	
95/64	6	7	6	7	6	C	D	
95/65	6	5	6	4	4	B	A	
95/66	4	4	6	4	2	B	A	
95/67	6	4	6	6	4	B	B	
95/68	3	3	4	4	3	A	A	
95/69	6	5	7	6	5	C	C	
95/70	6	7	7	7	6	C	C	
95/71	6	4	6	6	4	C	B	
95/72	4	5	6	5	3	B	A	
95/73	4	5	7	6	6	B	A	
95/74	6	5	6	6	5	B	C	
95/75	5	5	5	6	5	B	B	
95/76	6	4	6	4	3	C	A	
95/77	3	5	5	4	3	A	B	
95/78	4	3	4	4	1	A	A	
95/79	7	7	7	8	6	C	D	
95/80	7	6	6	6	5	B	D	
95/81	5	5	5	5	3	B	B	
95/82	5	5	6	5	5	B	B	
95/83	5	5	6	5	5	C	B	
95/84	3	4	4	3	3	B	B	
95/85	5	2	6	4	3	B	A	
95/86	5	5	6	7	4	C	C	
95/87	4	5	4	3	1	A	A	
95/88	7	5	7	7	6	B	C	
95/89	5	5	5	5	3	B	B	
95/90	4	5	5	5	5	C	C	
95/91	6	5	6	6	5	B	C	
95/92	5	7	6	5	5	C	B	
95/93	6	4	5	3	3	B	A	

95/94	6	5	5	5	3	B	A	
95/95	7	6	7	7	6	C	D	
95/96	7	7	7	6	6	C	C	
95/97	5	5	5	6	4	B	C	
95/98	7	5	6	4	5	B	A	
95/99	2	4	4	5	2	C	A	
95/100	5	5	5	5	3	B	A	
95/101	5	7	6	6	5	B	C	
95/102	4	4	4	4	3	B	A	
95/103	4	4	4	3	3	C	A	
95/104	4	4	5	5	4	B	B	
95/105	6	5	7	6	4	C	B	
95/106	5	5	5	5	3	C	B	
95/107	6	5	6	5	4	B	B	
95/108	4	6	6	5	5	C	C	
95/109	5	5	7	6	4	C	B	
95/110	6	6	6	7	5	C	B	
95/111	5	5	6	5	5	B	A	
95/112	7	6	7	8	5	C	B	
95/113	5	4	5	5	3	B	B	
95/114	4	5	5	6	4	B	A	
95/115	3	4	4	3	1	A	A	
95/116	6	7	7	6	6	C	D	
95/117	5	5	7	7	4	C	A	
95/118	6	4	5	5	4	A	B	
95/119	3	4	3	2	1	B	A	
95/120	5	3	5	6	3	A	A	
95/121	7	5	5	4	3	B	B	
95/122	5	7	6	7	5	B	B	
95/123	4	5	5	6	5	B	C	
95/124	4	4	5	5	4	B	B	
95/125	4	5	6	4	3	B	B	
95/126	5	5	5	5	4	B	B	
95/127	7	7	7	7	5	C	D	
95/128	7	5	6	6	4	C	C	
95/129	4	5	4	3	1	B	A	
95/130	6	5	6	5	3	C	B	
95/131	4	5	4	3	3	B	A	
95/132	6	5	7	7	5	C	C	
95/133	3	4	3	3	3	B	A	
95/134	5	5	6	5	5	B	A	
95/135	6	5	6	5	4	C	B	
95/136	6	5	7	7	6	B	C	
95/137	6	6	7	7	4	C	C	
95/138	7	5	6	7	6	B	B	
95/139	5	4	5	4	4	B	A	
95/140	7	6	7	7	5	C	C	
95/141	7	5	7	6	5	C	B	
95/142	7	6	7	6	4	B	C	
95/143	5	5	5	5	3	C	B	
95/144	6	5	7	7	5	B	B	
95/145	7	6	7	7	6	C	D	
95/146	6	7	7	7	5	C	C	
95/147	6	6	5	5	3	B	A	
95/148	5	5	5	5	3	C	B	

95/149	5	6	5	4	3	C	B	
95/150	4	4	4	4	3	B	B	
95/151	4	5	5	4	4	C	B	
95/152	6	5	6	5	5	B	C	
95/153	7	7	6	8	6	C	D	

APPENDIX 14

1996 INTAKE

RESULTS OF CAT

(LEVEL D, TAKEN IN 1996)

SS = Standard Age Score (Mean = 100)

	Verbal SS	Quantitative SS	Non-verbal SS	Average SS
96/1	85	78	78	80
96/2	118	107	112	112
96/3	103	114	107	108
96/4	98	85	94	92
96/5	97	83	99	93
96/6	126	117	117	120
96/7	96	91	106	98
96/8	105	105	103	104
96/9	109	101	101	104
96/10	130	130	115	125
96/11	84	70	78	77
96/12	108	108	119	112
96/13	124	115	111	117
96/14	104	89	88	94
96/15	95	92	96	94
96/16				
96/17	105	105	119	110
96/18	99	101	108	103
96/19	94	88	108	97
96/20	118	99	100	106
96/21				
96/22	90	94	93	92
96/23	87	88	111	95
96/24	109	95	95	100
96/25	109	100	110	106
96/26	98	101	108	102
96/27	98	82	94	91
96/28	130	112	112	118
96/29	88	91	75	85
96/30	95	92	97	95
96/31	98	87	94	93
96/32	91	94	94	93
96/33	98	98	104	100
96/34	106	113	115	111
96/35	98	84	99	94
96/36	84	78	88	83
96/37	104	105	109	106
96/38	88	87	111	95
96/39				
96/40	96	97	90	94
96/41				
96/42	121	110	108	113

96/43	100	102	107	103
96/44	86	90	96	91
96/45	82	72	84	79
96/46				
96/47	88	91	100	93
96/48	85	76	97	86
96/49	103	90	107	100
96/50	120	96	108	108
96/51				
96/52	83	85	84	84
96/53	110	95	104	103
96/54	84	71	82	79
96/55	119	117	114	117
96/56	115	118	112	115
96/57	82	85	100	89
96/58	103	100	112	105
96/59				
96/60	73	81	104	86
96/61	79	84	85	83
96/62	120	106	114	113
96/63	120	88	114	107
96/64	130	113	110	118
96/65	87	80	83	83
96/66	94	93	89	92
96/67	108	117	110	112
96/68	112	119	115	115
96/69	125	119	122	122
96/70	118	122	110	117
96/71	92	102	100	98
96/72	84	80	89	84
96/73	84	84	112	93
96/74	93	79	94	89
96/75	120	110	115	115
96/76	115	107	108	110
96/77	85	89	77	84
96/78	116	123	119	119
96/79	99	92	91	94
96/80	101	90	108	100
96/81	84	92	95	90
96/82	113	103	114	110
96/83	83	84	114	94
96/84	125	119	119	121
96/85	93	95	93	94
96/86	87	90	99	92
96/87	99	94	92	95
96/88	115	96	110	107
96/89	106	88	103	99
96/90	100	94	110	101
96/91	109	105	98	104
96/92	96	106	103	102
96/93				
96/94	101	104	99	101
96/95	79	80	92	84
96/96	122	109	121	117
96/97	104	117	113	111

96/98	105	97	117	106
96/99	74	84	80	79
96/100				
96/101	112	114	103	110
96/102				
96/103	107	113	103	108
96/104	124	125	112	120
96/105	88	81	100	90
96/106	102	90	107	100
96/107	87	96	104	96
96/108	116	109	129	118
96/109	115	108	110	111
96/110	94	94	94	94
96/111	108	110	115	111
96/112	83	105	101	96
96/113	84	79	77	80
96/114				
96/115	92	103	114	103
96/116	88	85	86	86
96/117	104	103	110	106
96/118	108	100	103	104
96/119	109	114	106	110
96/120	94	98	93	95
96/121	107	105	105	106
96/122	123	113	117	118
96/123	88	99	98	95
96/124	94	110	97	100
96/125	98	88	108	98
96/126	100	111	109	107
96/127				
96/128	86	80	78	81
96/129	79	102	89	90
96/130	125	110	101	112
96/131	89	94	86	90
96/132	98	74	88	87
96/133	98	84	96	93
96/134	87	72	84	81
96/135	126	104	117	116
96/136	86	80	81	82
96/137	105	107	110	107
96/138	109	98	107	105
96/139	93	101	87	94
96/140				
96/141				
96/142	91	83	101	92

APPENDIX 15

1996 INTAKE

KEY STAGE 2 RESULTS: CORE SUBJECTS (1996)

TEST AND TEACHER ASSESSMENT LEVELS

TA = Teacher assessment Eng = English Ma = Mathematics Sci = Science

	Eng Test	Eng TA	Ma Test	Ma TA	Sci Test	Sci TA	Test Average
96/1	3	4	3	3	3	4	3.00
96/2	4	4	4	4	4	4	4.00
96/3							
96/4	4	4	4	3	3	3	3.66
96/5	4	4	3	3	3	3	3.33
96/6	5	5	5	5	5	5	5.00
96/7	4	4	4	4	3	3	3.66
96/8	4	4	3	4	3	4	3.33
96/9	5	5	3	4	3	4	3.66
96/10	5	5	5	5	5	4	5.00
96/11							
96/12	4	4	5	5	5	5	4.66
96/13	5	6	5	6	5	5	5.00
96/14	4	3	3	3	4	3	3.66
96/15		4		4		4	
96/16							
96/17	4	5	4	5	4	5	4.00
96/18	4	4	4	4	4	4	4.00
96/19	4	4	4	4	4	4	4.00
96/20	4	4	4	4	4	4	4.00
96/21							
96/22	4	4	3	4	3	4	3.33
96/23	3	4	3	4	4	4	3.33
96/24							
96/25	5	4	4	4	4	4	4.33
96/26							
96/27	4	4	4	4	4	4	4.00
96/28	5	5	4	4	4	5	4.33
96/29							
96/30	4	4	3	3	3	4	3.33
96/31	4	3	3	3	4	3	3.66
96/32							
96/33	4	4	4	4	4	4	4.00
96/34	4	4	4	4	4	4	4.00
96/35	4	4	3	4	3	4	3.33
96/36		2		2	3	2	3.00
96/37	4	5	4	4	4	5	4.00
96/38	3	4	3	4	4	4	3.33
96/39	3	3	3	3	3	3	3.00
96/40							
96/41							
96/42	5	5	4	4	5	5	4.66

96/43							
96/44							
96/45	3	4		2	3	3	3.00
96/46							
96/47	3	3	3	3	3	3	3.00
96/48	3	3	3	3	3	3	3.00
96/49	4	4	4	4	4	4	4.00
96/50	4	5	4	4	4	4	4.00
96/51							
96/52	3	3	3	3	3	3	3.00
96/53	4	4	3	4	4	4	3.66
96/54	3	3	3	3	3	3	3.00
96/55	5	5	5	5	5	4	5.00
96/56	5	5	4	5		5	4.50
96/57	3	3	3	3	3	3	3.00
96/58							
96/59							
96/60							
96/61	3	3	3	3	3	3	3.00
96/62	5	5	5	5	4	4	4.66
96/63	4	4	4	4	5	4	4.33
96/64	5	6	4	4	5		4.66
96/65	3	3	3	3	3	3	3.00
96/66	4	4	4	4	3	4	3.66
96/67	4	4	4	5	5	5	4.33
96/68	5	3	5	4	4	4	4.66
96/69	5	5	5	5	4	5	4.66
96/70	5	5	5	5	5	5	5.00
96/71	4	4	4	4	3	4	3.66
96/72							
96/73	3	3	3	3	3	3	3.00
96/74	3	4	3	3	3	3	3.00
96/75	5	5	5	5	5	5	5.00
96/76	4	4	4	4	4	4	4.00
96/77	4	4	3	3	3	4	3.33
96/78	5	5	5	5	5	5	5.00
96/79	4	4	3	3	4	4	3.66
96/80	3	3	4	3	4	4	3.66
96/81							
96/82	5	5	4	4	4	4	4.33
96/83							
96/84							
96/85	3	3	4	3		3	3.50
96/86	3	3	4	3		3	3.50
96/87	4	4	3	3	4	4	3.66
96/88	4	4	4	4	4	4	4.00
96/89	4	4	3	4	4	4	3.66
96/90	4	4	4	4	4	3	4.00
96/91	4	4	3	4	4	4	3.66
96/92	3	4	4	4	4	4	3.66
96/93							
96/94	4	4	4	4	4	4	4.00
96/95	3	3	3	3	3	3	3.00
96/96	4	5	5	5	4	4	4.66
96/97	4	4	5	4	4	4	4.33

96/98	4	4	4	4	4	4	4.00
96/99	4	4	4	4	3	3	3.66
96/100	3	3	3	3	3	3	3.00
96/101	4	5	5	5	4	5	4.33
96/102							
96/103	5	4	4	4	4	4	4.33
96/104	5	5	5	5	5	5	5.00
96/105	4	4	3	3	4	4	3.66
96/106	4	3	4	4	4	3	4.00
96/107	4	4	4	4	4	4	4.00
96/108	4	5	4	5	4	4	4.00
96/109	4	5	4	4	4	5	4.00
96/110	3	4	4	4	4	4	3.66
96/111	3	3	3	3	3	3	3.00
96/112	3	3	3	3	3	3	3.00
96/113	3	4	3	3	3	3	3.00
96/114							
96/115	3	4	3	3	3	4	3.00
96/116	5	5	4	4	4	4	4.33
96/117							
96/118	5	5	4	4	4	4	4.33
96/119	5	5	5	5	5	5	5.00
96/120	3	3	3	3	3	3	3.00
96/121	4	5	4	4	4	4	4.00
96/122							
96/123	3	3	3	3	3	4	3.00
96/124	4	4	3	4	3	3	3.33
96/125		4		4		4	
96/126	4	4	4	4	4	4	4.00
96/127							
96/128	4	3	2	3	3	3	3.00
96/129	3	4	4	4	3	4	3.33
96/130	5	5	4	5	5	5	4.66
96/131	3	4	3	3	3	3	3.00
96/132	4	4	3	3	4	4	3.66
96/133	3	4	4	4	3	3	3.33
96/134	3	3	3	3	3	3	3.00
96/135	5	5	4	4	5	5	4.66
96/136	3	3		2	3	2	3.00
96/137	4	5	4	5	4	4	4.00
96/138	4	4	4	4	4	4	4.00
96/139	3	3		3		3	3.00
96/140							
96/141							
96/142	3	3	3	3	3	3	3.00

APPENDIX 16

1996 INTAKE

KEY STAGE 3 RESULTS: CORE SUBJECTS (1999)

TA = Teacher assessment Eng = English Ma = Mathematics Sci = Science

	Eng TA	Eng Test	Ma TA	Ma Test	Sci TA	Sci Test	Test Average
96/1	5	5	4	4	3	4	4.33
96/2	6	6	6	6	6	6	6.00
96/3	6	5	7	7	5		6.00
96/4	5	5	4	4	4	4	4.33
96/5	3	3	3	3	4	3	3.00
96/6	7	7	8	8	7	7	7.33
96/7	6	5	6	7	5	6	6.00
96/8	5	5	5	5	4	4	4.66
96/9	6	6	5	5	4	4	5.00
96/10	8	8			7	7	7.50
96/11	5		3	3	3	3	3.00
96/12	6	5	7	7	6	6	6.00
96/13	8	8	8	7	7	7	7.33
96/14	6	5	5	5	4	4	4.66
96/15	4	4	4	5	5	5	4.66
96/16	6	5	5	6	5	5	5.33
96/17	6	5	7	7	5	6	6.00
96/18	6	6	6	6	5	6	6.00
96/19	5	5	5	5	5	5	5.00
96/20	6	6	6	6	5	6	6.00
96/21							
96/22	5	5	5	5	5	5	5.00
96/23	5	5	5	5	4	4	4.66
96/24	6	6	5	6	6	6	6.00
96/25	5	5	5	6	5	6	5.66
96/26	5	5	5	5	5	5	5.00
96/27	4	5	5	5	5	5	5.00
96/28	7	7	7	7	6	7	7.00
96/29	5	5	4	4	4	5	4.66
96/30	6	6	5	5	5	5	5.33
96/31	5	5	4	4	5	5	4.66
96/32	5	5	4	4	4	4	4.33
96/33	5	5	5	5	4	4	4.66
96/34	6	5	6	6	5	6	5.66
96/35	6	6	4	4	4	5	5.00
96/36	3		3		3	3	3.00
96/37	6	5	5	5	5	5	5.00
96/38	4	5	4	4	4	4	4.33
96/39	4	4	3	3	3	4	3.66
96/40	5	5	5	5	5	5	5.00
96/41	4	4	3	3	4	4	3.66
96/42	6	6	6	6	5	6	6.00
96/43	5	5	6	5	4	5	5.00
96/44	3		4	4	4	4	4.00

96/45	5	5	4	4	5	5	4.66
96/46	5	5	5	5	4	4	4.66
96/47	5	5	5	5	4	5	5.00
96/48	4	5	4	4	4	4	4.33
96/49	5	5	5	6	4	5	5.33
96/50	6	6	5	5	5		5.50
96/51	3	4	4	4	4	4	4.00
96/52	4	3	5	5	4	4	4.00
96/53	7	7	6	6	7	6	6.33
96/54	4	3	3	3	3	3	3.00
96/55	7	7	8	7	7	7	7.00
96/56	6	7	7	7	7	6	6.66
96/57	4	4	4	4	4	4	4.00
96/58	5	5	5	6	5	6	5.66
96/59	6	5	6	6	6	6	5.66
96/60	5	5	4	5	5	5	5.00
96/61	4	4	4	4	4	4	4.00
96/62	7	7	7	7	7	7	7.00
96/63	6	6	7	7	6	7	6.66
96/64	7	7	7	7	6	6	6.66
96/65	4	4	4	4	3	4	4.00
96/66	5	5	6	5	5	5	5.00
96/67	6	5	7	7	7	7	6.33
96/68	6	6	7	7	5	6	6.33
96/69	6	6	7	7	7	7	6.66
96/70	7	7	8	8	7	7	7.33
96/71	4	4	4	5	3	3	4.00
96/72	5	5	4	4	4	4	4.33
96/73	4	4	4	4	4	4	4.00
96/74	5	5	4	5	5	4	4.66
96/75	7	8	7	7	5	6	7.00
96/76	6	6	6	6	5	5	5.66
96/77	4	5	5		4	4	4.50
96/78	6	7	7	7	6	6	6.66
96/79	5	5	5	5	5	5	5.00
96/80	5	5	6	6	5	5	5.33
96/81	5	4	5		5	4	4.00
96/82	5	6	6	6	6	6	6.00
96/83	4	4	4	4	4	4	4.00
96/84	7	6	8	8	5	6	6.66
96/85	4	5	5	5	5	4	4.66
96/86	5	4	4	4	4	5	4.33
96/87	4	5	4	4	4	5	4.66
96/88	6	5	5	6	5		5.50
96/89	5	5	4	5	5	5	5.00
96/90	6	5	6	6	5	6	5.66
96/91	5	5	5	5	4	4	4.66
96/92	4	5	5	5	5	5	5.00
96/93	4		4	4	4	4	4.00
96/94	6	6	5	5	5	5	5.33
96/95	4	5	4	4	4	4	4.33
96/96	6	6	8	8	7	7	7.00
96/97	6	6	7	7	5	6	6.33
96/98	6	6	6	6	5	5	5.66
96/99	4	5	4	4	4	4	4.33

96/100	4	4	4	4	4	4	4.00
96/101	7	7	6	6	5	6	6.33
96/102	3	4	4	4	4	4	4.00
96/103	7	7	6	6	5	5	6.00
96/104	7	7	7	7	6	6	6.66
96/105	5	5	4	5	5	5	5.00
96/106	5	4	4	4	4	4	4.00
96/107	5	5	5	5	5	4	4.66
96/108	7	7	7	7	6	6	6.66
96/109	7	5	6	6	6	6	5.66
96/110	4	4	4	4	4	3	3.66
96/111	6	6	5	6	6	6	6.00
96/112	5	5	6	5	4	4	4.66
96/113	5	5	4	4	4	5	4.66
96/114	5	3	5	6	5	5	4.66
96/115	5	5	6	5	4	5	5.00
96/116	5	5	4	4	4	4	4.33
96/117	6	5	5	6	4	5	5.33
96/118	6	6	5	6	6	5	5.66
96/119	7	6	7	7	6	6	6.33
96/120	5	4	6	5	4	4	4.33
96/121	6	5	6	6	4	5	5.33
96/122	7	6	7	7	7	6	6.33
96/123	5	5	5	4	4	4	4.33
96/124	5	5	5	6	4	4	5.00
96/125	6	5	4	5	5	5	5.00
96/126	7	6	6	7	5	6	6.33
96/127	6	5	6	6	5	5	5.33
96/128	4	5	3	4	4	4	4.33
96/129	5	5	5	5	5	5	5.00
96/130	7	7	7	7	7	7	7.00
96/131	5	5	4	4	4	4	4.33
96/132	5	5	4	4	4	5	4.66
96/133	4	4	5	5	4	4	4.33
96/134	4	4	3	3	4	4	3.66
96/135	8	7	7	7	7	7	7.00
96/136	5	4	3	3	4	4	3.66
96/137	5	5	5	6	5	5	5.33
96/138	5	5	5	6	5	5	5.33
96/139	4	5	5	5	4	4	4.66
96/140	6	7	6	6	6	7	6.66
96/141	5	5	5	5	4	5	5.00
96/142	4	5	5	5	4	4	4.66

APPENDIX 17

1996 INTAKE

KEY STAGE 3 RESULTS: NON CORE SUBJECTS (1999)

TEACHER ASSESSMENT LEVELS

DT = Design Technology Hi = History Geo = Geography Mu = Music

ICT = Information and Communications Technology

PE = Physical Education MFL = Modern Foreign Languages

	Hi	Geo	DT	ICT	MFL	Art	Mu	PE
96/1	4	3	4	4	4	B	B	B
96/2	7	7	6	6	5	C	C	B
96/3	5	6	5	6	5	C	A	B
96/4	4	4	3	5	4	C	B	B
96/5	3	4	3	4	3	C	B	B
96/6	8	7	7	6	6	C	C	B
96/7	5	5	5	5	4	C	C	B
96/8	5	4	4	5	4	C	B	B
96/9	5	4	4	6	4	B	B	B
96/10	8	7	7	7	6	C	C	B
96/11	3	4	3	4	2	C	A	B
96/12	5	6	5	6	4	C	C	B
96/13	8	7	7	6	6	C	D	B
96/14	5	5	5	5	4	B	B	B
96/15	5	5	6	5	5	C	B	B
96/16	6	4	4	5	5	B	D	C
96/17	6	6	6	6	5	C	C	B
96/18	7	7	8	5	4	C	C	B
96/19	5	4	5	5	4	C	B	B
96/20	7	7	6	6	4	C	C	B
96/21								
96/22	5	5	6	5	4	C	C	C
96/23	4	4	5	4	4	C	A	B
96/24	5	5	7	5	5	B	B	B
96/25	7	5	5	5	4	C	A	B
96/26	5	4	5	5	4	C	B	B
96/27	5	4	4	4	4	B	B	B
96/28	7	7	7	6	5	C	C	B
96/29	5	5	5	4	5	B	A	B
96/30	6	5	4	5	4	C	C	B
96/31	4	5	5	5	4	C	A	B
96/32	5	4	6	4	4	C	A	B
96/33	4	4	5	5	2	B	A	B
96/34	5	5	5	6	5	C	B	B
96/35	5	4	5	5	4	C	B	B
96/36	4	2	4	4	2	B	A	A
96/37	5	5	6	5	4	C	B	B
96/38	4	3	3	4	4	B	A	B

96/39	4	4	4	3	3	C	A	A
96/40	5	5	5	5	5	C	B	B
96/41	4	3	3	3	4	C	A	B
96/42	7	6	6	6	5	C	C	B
96/43	6	5	6	5	5	C	C	B
96/44								
96/45	5	6	6	4	5	C	B	B
96/46	4	4	5	4	4	B	A	B
96/47	5	5	5	4	4	B	B	B
96/48	4	4	7	5	4	C	B	B
96/49	5	5	5	5	4	C	B	B
96/50	6	4	4	5	4	C	B	B
96/51	4	3	3	3	3	B	A	B
96/52	4	3	4	5	4	C	C	B
96/53	7	7	6	6	5	B	D	B
96/54	3	3	5	4	2	B	A	B
96/55	8	7	8	6	5	C	D	B
96/56	7	7	7	6	5	C	D	B
96/57	5	5	5	3	3	C	B	B
96/58	7	5	5	5	4	C	B	B
96/59	6	7	6	3	5	C	B	B
96/60	4	4	4	3	2	C	A	B
96/61	4	4	4	4	3	B	A	B
96/62	8	7	7	6	5	C	D	C
96/63	6	7	7	6	5	C	B	B
96/64	7	7	7	6	5	B	C	B
96/65	4	4	4	3	3	C	A	B
96/66	5	6	6	5	5	C	A	B
96/67	6	6	5	6	5	B	B	B
96/68	6	6	6	6	5	B	C	B
96/69	7	6	7	6	5	C	D	B
96/70	7	7	8	6	5	C	C	B
96/71	4	4	4	4	4	C	B	B
96/72	5	5	5	4	3	B	A	B
96/73	5	4	6	4	4	C	C	B
96/74	5	5	6	4	4	C	B	B
96/75	7	7	7	6	5	C	C	B
96/76	7	6	5	6	4	C	C	B
96/77	5	4	5	5	3	C	B	B
96/78	7	7	7	6	6	B	D	C
96/79	5	5	4	5	5	C	A	B
96/80	7	6	4	5	4	C	A	B
96/81	5	5	5	5	4	B	A	B
96/82								
96/83	4	4	4	3	4	C	B	B
96/84	7	6	5	6	5	B	D	B
96/85	4	5	4	5	3	B	A	B
96/86	3	4	4	4	4	B	A	B
96/87	5	5	4	4	5	C	A	B
96/88	7	6	5	5	3	C	B	A
96/89	5	4	5	5	4	C	B	B
96/90	6	6	6	6	5	C	B	B
96/91	5	4	5	4	5	C	A	B
96/92	4	4	5	5	3	B	B	B
96/93	4	3	4	3	3	B	B	B

96/94	6	4	6	5	4	C	C	B
96/95	4	4	5	4	4	C	B	B
96/96	7	7	7	6	5	C	C	B
96/97	6	6	7	6	5	C	C	B
96/98	6	7	6	5	5	C	B	B
96/99	4	4	5	4	4	B	B	B
96/100	6	5	5	5	5	C	A	B
96/101	7	6	6	6	5	C	C	B
96/102	5	5	5	3	5	B	B	B
96/103	7	6	6	5	5	C	D	C
96/104	7	7	6	6	5	C	D	B
96/105	5	5	5	4	4	C	B	B
96/106	5	4	4	4	2	C	A	A
96/107	5	5	6	4	5	C	B	B
96/108	7	6	7	6	5	C	C	B
96/109	8	7	6	6	4	C	D	B
96/110	4	3	6	5	2	B	B	B
96/111	6	6	6	6	5	C	C	B
96/112	5	4	5	4	3	B	B	B
96/113	5	5	5	4	4	C	A	B
96/114	5	4	4	3	2	C	A	B
96/115	4	5	5	4	5	C	B	B
96/116	4	4	4	3	3	C	A	B
96/117	6	5	5	5	4	C	D	B
96/118	6	6	7	6	4	C	D	B
96/119	6	7	7	7	5	C	C	B
96/120	4	4	4	4	4	B	A	B
96/121	6	6	5	5	5	B	C	B
96/122	7	6	7	6	5	C	C	B
96/123	4	5	5	4	5	C	B	B
96/124	5	4	5	5	4	B	B	B
96/125	6	5	5	4	4	C	C	B
96/126	6	7	6	5	5	B	B	B
96/127	6	6	5	5	8	C	B	B
96/128	4	4	5	4	3	C	A	B
96/129	6	6	7	4	5	C	A	B
96/130	6	7	6	6	5	C	C	B
96/131	4	3	5	4	4	C	A	B
96/132	4	4	5	4	3	C	C	B
96/133	4	4	3	4	3	B	B	B
96/134	4	4	4	4	3	C	A	B
96/135	6	7	7	6	5	C	D	B
96/136	4	4	5	4	5	B	B	B
96/137	5	4	7	5	4	C	C	B
96/138	6	5	6	5	4	C	B	C
96/139	4	5	4	4	4	C	B	B
96/140	7	7	6	6	5	C	B	B
96/141	4	4	4	5	4	B	B	B
96/142	4	3	5	5	4	C	B	B

APPENDIX 18

1997 INTAKE

RESULTS OF CAT

(LEVEL D, TAKEN IN 1997)

SS = Standard Age Score (Mean = 100)

	Verbal SS	Quantitative SS	Non-verbal SS	Average SS
97/1				
97/2	76	83	74	78
97/3	118	120	115	118
97/4	98	92	105	98
97/5	99	119	103	107
97/6	98	99	91	96
97/7	98	99	97	98
97/8	103	109	105	106
97/9	121	116	109	115
97/10	112	117	114	114
97/11	82	83	86	84
97/12	115	123	111	116
97/13	103	99	103	102
97/14	95	94	96	95
97/15	100	93	96	96
97/16	130	115	125	123
97/17	109	99	114	107
97/18	92	102	101	98
97/19	95	100	91	95
97/20	114	88	104	102
97/21	107	95	106	103
97/22	104	87	110	100
97/23	96	100	89	95
97/24	91	98	89	93
97/25	81	86	80	82
97/26	95	89	102	95
97/27	117	114	117	116
97/28				
97/29	79	78	72	76
97/30	85	103	92	93
97/31	85	91	89	88
97/32	101	95	97	98
97/33				
97/34	102	94	96	97
97/35	99	103	112	105
97/36	79	82	99	87
97/37	75	96	99	90
97/38	81	83	85	83
97/39	104	94	96	98
97/40				
97/41	92	113	114	106
97/42	87	87	86	87

97/43	78	70	79	76
97/44	102	100	88	97
97/45	97	110	95	101
97/46	100	109	84	98
97/47	126	110	109	115
97/48	120	114	104	113
97/49	74	81	104	86
97/50	87	89	94	90
97/51	105	97	113	105
97/52	95	93	107	98
97/53	89	89	88	89
97/54	101	94	105	100
97/55	100	91	92	94
97/56	98	89	96	94
97/57	88	96	98	94
97/58	128	117	126	124
97/59	105	82	91	93
97/60				
97/61	107	112	109	109
97/62	113	103	103	106
97/63				
97/64	98	88	105	97
97/65	94	95	106	98
97/66	98	110	106	105
97/67	77	83	99	86
97/68	96	96	107	100
97/69	100	90	101	97
97/70	116	100	97	104
97/71	129	97	109	112
97/72	89	101	93	94
97/73	92	92	99	94
97/74	90	88	97	92
97/75	112	93	117	107
97/76				
97/77	109	114	109	111
97/78	108	104	112	108
97/79	82	90	96	89
97/80	93	92	86	90
97/81	107	94	109	103
97/82	106	107	110	108
97/83	92	104	108	101
97/84				
97/85	104	96	101	100
97/86	116	99	100	105
97/87	107	115	117	113
97/88	108	115	117	113
97/89	96	89	93	93
97/90	106	96	105	102
97/91	103	113	107	108
97/92	109	115	97	107
97/93	109	104	106	106
97/94	89	87	97	91
97/95	125	105	122	117
97/96	110	111	107	109
97/97	81	70	70	73

97/98	72	75	75	74
97/99	72	70	71	0
97/100	115	105	117	112
97/101	98	83	85	89
97/102	101	93	107	100
97/103	114	116	122	117
97/104				
97/105	89	96	100	95
97/106	99	87	102	96
97/107	89	91	91	90
97/108	89	104	82	92
97/109	84	84	84	84
97/110	83	80	88	84
97/111	109	97	108	105
97/112	91	79	92	87
97/113	95	92	104	97
97/114	97	82	89	89
97/115	102	96	101	100
97/116	130	108	105	115
97/117	123	101	110	111
97/118	90	82	82	85
97/119				
97/120	85	85	79	83
97/121	84	81	74	80
97/122	126	111	122	120
97/123				
97/124	111	116	106	111
97/125	98	95	98	97
97/126	90	92	86	89
97/127				
97/128	127	115	108	117
97/129	84	72	76	77
97/130	96	101	113	103
97/131	126	114	117	119
97/132	98	99	99	99
97/133	103	91	106	100
97/134	91	92	89	91
97/135	98	94	97	96
97/136	84	75	74	78
97/137	122	99	107	109
97/138	101	104	106	104
97/139	98	94	106	99
97/140	117	130	117	122
97/141	103	90	97	97
97/142	91	86	100	92
97/143	91	94	106	97
97/144	70	82	101	84
97/145	94	88	109	97
97/146	111	97	103	104
97/147	91	88	117	99
97/148	99	89	100	96
97/149	98	88	84	90
97/150	73	70	70	70
97/151	85	94	72	84
97/152	87	82	96	88

97/153	92	109	110	104
97/154	100	89	88	92
97/155	130	114	125	123
97/156	118	91	108	106
97/157	125	96	112	111
97/158	121	116	119	119
97/159	117	100	98	105
97/160	105	94	86	95
97/161	76	76	74	75
97/162	95	95	96	95
97/163	76	78	84	79
97/164	71	77	70	72
97/165	75	81	80	79
97/166	108	109	108	108

APPENDIX 19

1997 INTAKE

KEY STAGE 2 RESULTS: CORE SUBJECTS (1997)

TA = Teacher Assessment Eng = English Ma = Mathematics Sci= Science

	Eng Test	Eng TA	Ma Test	Ma TA	Sci Test	Sci TA	Test average
97/1	4		4		4		4.00
97/2	3	3		2	3	3	3.00
97/3	5	5	5	5	5	5	5.00
97/4							
97/5	4	4	4	4	3	3	3.66
97/6	4	4	3	4	4	4	3.66
97/7	4	4	4	4	3	4	3.66
97/8	4	4	4	5	4	4	4.00
97/9	5	5	5	5	4	5	4.66
97/10	3		4		3		3.33
97/11	3	3	4	4	3	3	3.33
97/12							
97/13	4	4	4	4	3	4	3.66
97/14	4	4	4	4	4	4	4.00
97/15	3	3	3	3	4	3	3.33
97/16	5	5	5	5	5	5	5.00
97/17	4	5	4	4	5	5	4.33
97/18	4		3		4		3.66
97/19	4	4	4	4	4	4	4.00
97/20	4	4	4	4	4	5	4.00
97/21	4	4	4	3	4	3	4.00
97/22	4	4	4	4	4	4	4.00
97/23	4	4	4	5	4	4	4.00
97/24	4		3		4		3.66
97/25	3	3	3	3	3	3	3.00
97/26	4	4	4	4	4	4	4.00
97/27	4		4		5		4.33
97/28	4		3		4		3.66
97/29	3		3		3		3.00
97/30	3	3	3	3		4	3.00
97/31	3	3	3	3	3	3	3.00
97/32		4		4		4	
97/33	4		4		4		4.00
97/34	4		4		4		4.00
97/35	4	5	4	4	4	4	4.00
97/36	3	3	3	2	3	3	3.00
97/37	3	3	3	3	3	2	3.00
97/38	3	3	3	3	3	3	3.00
97/39	5	5	4	4	4	4	4.33
97/40	4		5		4		4.33
97/41	4	5	5	5	4	4	4.33
97/42	4	3	3	3	4	4	3.66
97/43	3	4	3	3	3	4	3.00
97/44	4	4	4	4	4	4	4.00

97/45	4	4	4	4	4	4	4.00
97/46	4		4		4		4.00
97/47	5	5	5	5	5	5	5.00
97/48	5	5	5	5	5	5	5.00
97/49	3	3	3	3	3	3	3.00
97/50	3	3	4	4	3	4	3.33
97/51	4	4	4	4	4	3	4.00
97/52	4	4	3	4	3	4	3.33
97/53	3	4	3	4	3	4	3.00
97/54	4	4	4	4	4	4	4.00
97/55		4		3		4	
97/56	4	4	3	3	3	3	3.33
97/57	3	3	4	3	3	3	3.33
97/58	5	5	5	5	5	5	5.00
97/59	4	4	4	4	5	5	4.33
97/60	4		3		3		3.33
97/61	4	4	4	4	5	5	4.33
97/62	4	5	4	4	5	5	4.33
97/63	4		5		4		4.33
97/64		4		4		4	
97/65	4		4		4		4.00
97/66	4		5		4		4.33
97/67	4	3	3	3	3	3	3.33
97/68							
97/69	4	5	4	4	4	4	4.00
97/70	4	4	4	4	4	4	4.00
97/71	5	5	4	4	4	4	4.33
97/72	3	4	4	4	4	4	3.66
97/73	3	3	4	4	4	4	3.66
97/74	3	3	3	3	3	3	3.00
97/75	5	4	4	4	4	4	4.33
97/76	3	3	4	4	4	4	3.66
97/77	4	5	4	4	4	4	4.00
97/78	4		4		4		4.00
97/79	3	3	3	3	3	3	3.00
97/80	4	4	4	3	4	4	4.00
97/81		5		5		5	
97/82	5	5	4	4	4	4	4.33
97/83	4	4	4	4	3	3	3.66
97/84	4		3		3		3.33
97/85	4	3	4	4	4	4	4.00
97/86							
97/87	4		5		4		4.33
97/88	4		5		4		4.33
97/89	4	4	4	4	4	4	4.00
97/90	4	4	3	3	4	4	3.66
97/91	5	5	4	4	4	4	4.33
97/92	4	4	4	4	4	4	4.00
97/93	4	4	4	5	4	5	4.00
97/94	4	4	3	4	4	4	3.66
97/95	4	4	4	5	4	5	4.00
97/96	4	5	4	5	4	4	4.00
97/97		2		1	3	2	3.00
97/98							
97/99							

97/100	5	5	5	5	5	5	5.00
97/101	4	4	4	4	4	4	4.00
97/102	5	5	4	4	4	4	4.33
97/103	4	5	5	5	5	5	4.33
97/104	4		3		3		3.33
97/105	4	4	4	4	4	4	4.00
97/106							
97/107	4		3		3		3.33
97/108	4	4	4	4	3	3	3.66
97/109	3	3	3	3	4	3	3.33
97/110	3	3		2	3	2	3.00
97/111	4	4	4	4	4	4	4.00
97/112	4	4	3	3	3	3	3.33
97/113	4		3		4		3.66
97/114	4	4	3	3	4	3	3.66
97/115	4	4	4	3	4	4	4.00
97/116	5	5	4	4	4	4	4.33
97/117							
97/118	3	3	3	3	3	3	3.00
97/119	4	4	4	4	4	4	4.00
97/120	4	4	3	3	4	4	3.66
97/121	3	4	3	3	3	4	3.00
97/122	5	5	5	5	5	5	5.00
97/123	3	4	4	4		4	3.50
97/124	4		4		5		4.33
97/125	3	4	3	4	3	4	3.00
97/126	3	3	3	4	3	2	3.00
97/127							
97/128	5	5	5	5	5	5	5.00
97/129	3		3		4		3.33
97/130	3	4	4	4	4	4	3.66
97/131	5		5		4		4.66
97/132	4	4	4	4	4	4	4.00
97/133	4	4	4	4	4	4	4.00
97/134	4	4	3	4	4	4	3.66
97/135	5	4	4	4	4	4	4.33
97/136	3				3		3.00
97/137	4	4	4	4	5	5	4.33
97/138	5	3	4	4	4	3	4.33
97/139	4	4	4	4	4	4	4.00
97/140							
97/141	4		4		4		4.00
97/142	3	3		3		3	3.00
97/143	3	4	4	4	4	4	3.66
97/144	3	3		3	3	3	3.00
97/145	4	3	4	3	4	3	4.00
97/146	5	4	4	4	5	4	4.66
97/147	4	4	3	4	3	4	3.33
97/148		3		3		4	
97/149	4	4	3	3	3	3	3.33
97/150	2	3	2	3	3	3	2.33
97/151	3	2	3	3		2	3.00
97/152	3	3	3	3	3	3	3.00
97/153	3	4	4	4	4	4	3.66
97/154	4		4		4		4.00

97/155	5	5	5	5	5	5	5.00
97/156	4		4		5		4.33
97/157							
97/158	5	5	4	5	4	5	4.33
97/159	4		4		5		4.33
97/160	4	4	3	3	4	4	3.66
97/161	3	3	3	3		3	3.00
97/162	4	4	4	4	4	4	4.00
97/163		3	2	3	3	3	2.50
97/164		2		2	3	3	3.00
97/165	3	3		3	3	3	3.00
97/166	4	4	4	4	4	4	4.00

APPENDIX 20

1997 INTAKE

KEY STAGE 3 RESULTS: CORE SUBJECTS (2000)

TA = Teacher assessment Eng = English Ma = Mathematics Sci = Science

	Eng Test	Eng TA	Ma Test	Ma TA	Sci Test	Sci TA	Test Average
97/1		3		4			
97/2	4	4	3	4		4	3.50
97/3	6	7	8	8	7	7	7.00
97/4	5	7	6	5	6	6	5.66
97/5	5	5	6	6	5	5	5.33
97/6	5	6	6	5	6	6	5.66
97/7	6	5	6	5	4	4	5.33
97/8	6	5	7	6	5	4	6.00
97/9	6	6	7	7	6	7	6.33
97/10	6	7	7	7	6	7	6.33
97/11	5	4	5	5	3	4	4.33
97/12	6	6	7	7	6	6	6.33
97/13	6	7	6	6	5	5	5.66
97/14	5	5	5	5	4	4	4.66
97/15	5	4	4	4	4	4	4.33
97/16	6	7	7	7	7	6	6.66
97/17	5	5	7	7	7	6	6.33
97/18	6	4	5	5	4	4	5.00
97/19	5	5	5	5	5	6	5.00
97/20	5	5	6	5	6	6	5.66
97/21	4	5	5	5	5	5	4.33
97/22	6	5	5	5	5	4	5.33
97/23	5	5	6	6	6	6	5.66
97/24	6	5	5	5	4	4	5.00
97/25	5	4	4	4	3	4	4.00
97/26	6	6	5	5	6	5	5.66
97/27	6	7	7	7	6	6	6.33
97/28	6	6	4	4		4	5.00
97/29		4	4	4	4	4	4.00
97/30	4	4	4	4	4	3	4.00
97/31	4	4	4	4	3	3	3.66
97/32	4	5	6	6	5	5	5.00
97/33	5	5	4	5	5	5	4.66
97/34	5	5	5	5	4	4	4.66
97/35	6	6	6	6	5	6	5.66
97/36	4	4	4	4	4	4	4.00
97/37	5	5	5	5	4	4	4.66
97/38	4	4	4	4		4	4.00
97/39		5		5		4	
97/40	7	7	7	7	6	7	6.66
97/41	6	5	7	8	6	6	6.33
97/42	5	4	4	4	3	3	4.00
97/43	4	4	4	4	4	4	4.00
97/44	6	4	5	5	5	4	5.33

97/45	5	5	6	6	5	4	5.33
97/46	5	5	5	4	5	4	5.00
97/47	8	7	7	7	7	7	7.33
97/48	8	7	7	7	7	6	7.33
97/49	4	4	4	4	4	3	4.00
97/50	5	4	5	5	5	4	5.00
97/51	5	6	6	6	5	6	5.33
97/52	4	5	5	5	4	4	4.33
97/53	4	4	4	4	4	3	4.00
97/54	5	5	6	5	5	6	5.33
97/55	6	6	4	4	5	4	5.00
97/56	4	4	4	4	4	3	4.00
97/57		3	5	5	4	4	4.50
97/58	6	7	8	8	7	6	7.00
97/59	6	5	5	5	6	5	5.66
97/60	4	5	4	4	4	5	4.00
97/61	6	7	7	6	7	6	6.66
97/62	6	6	7	6	7	6	6.66
97/63	6	6	6	5	5	4	5.66
97/64	5	5	4	4	5	4	4.66
97/65	5	5	6	5	5	4	5.33
97/66	6	6	7	6	5	6	6.00
97/67	4	4	4	4	4	4	4.00
97/68	5	5	5	5	5	4	5.00
97/69	6	6	5	5	5	5	5.66
97/70		5		5		5	
97/71	7	7	5	5	6	6	6.00
97/72	5	5	5	5		4	5.00
97/73	5	5	5	5		5	5.00
97/74	4	4	4	4	4	4	4.00
97/75	6	6	5	5	6	6	5.66
97/76	4	4	5	5	4	4	4.33
97/77	6	6	7	7	6	5	6.33
97/78	7	5	6	6	6	5	6.33
97/79	5	4	4	4	4	4	4.33
97/80	5	4	4	4	5	4	4.66
97/81	6	6	6	6	6	6	6.00
97/82	6	6	7	7	6	6	6.33
97/83	5	4	5	5	4	4	4.66
97/84	5	5	4	5	4	4	4.33
97/85	5	5	5	6	5	4	5.00
97/86	7	7	6	6	6	6	6.33
97/87	6	6	7	7	6	7	6.33
97/88	5	5	7	6	5	5	5.66
97/89	5	6	6	5	5	4	5.33
97/90	6	6	5	5	5	5	5.33
97/91	7	6	7	7	4	5	6.00
97/92	7	6	6	6	6	6	6.33
97/93	6	7	7	7	7	7	6.66
97/94	6	6	5	4	5	4	5.33
97/95	6	5	6	6	6	6	6.00
97/96	5	5	7	7	6	5	6.00
97/97	4	4	2	2	2	4	2.66
97/98	5	4	3	3	3	3	3.66
97/99	4	4	3	3	3	3	3.33

97/100	7	7	6	6	6	6	6.33
97/101	5	4	5	5	5	4	5.00
97/102	4	5	5	5	4	5	4.33
97/103	8	7	7	7	6	5	7.00
97/104	5	6	7	6	5	4	5.66
97/105	5	5	5	5	4	4	4.66
97/106	5	5	5	5	5	5	5.00
97/107	4	4	4	4	4	4	4.00
97/108	5	5	5	5	5	5	5.00
97/109	4	4	4	4	4	4	4.00
97/110	3	4	4	4		3	4.00
97/111	5	6	5	5	6	5	5.33
97/112	5	4	5	5	4	4	4.66
97/113	5	6	5	5	5	4	5.00
97/114	6	5	4	4	4	4	4.66
97/115	5	4	5	5	4	5	4.66
97/116	8	7	6	7	6	6	6.66
97/117	5	7	7	7	7	7	6.33
97/118	5	4	4	4	4	3	4.33
97/119	5	5	5	5	5	5	5.00
97/120	5	4	4	4	3	4	4.00
97/121	5	4	3	3	3	3	3.66
97/122	7	7	8	8	7	7	7.33
97/123		5		5		4	
97/124	6	6	7	7	7	6	6.66
97/125	6	6	6	5	5	4	5.66
97/126	5	4	4	4	4	4	4.33
97/127	6	6	6	7	6	6	6.00
97/128	6	7	7	7	7	6	6.66
97/129	4	4	3	3	3	3	3.33
97/130	4	5	6	6	6	6	5.33
97/131	7	7	7	7	6	6	6.66
97/132	6	6	6	6	6	5	6.00
97/133	5	5	5	6	5	5	5.00
97/134	5	4	4	4	4	4	4.33
97/135	5	6	5	4	5	4	5.00
97/136	5	4	3	3	3	3	3.66
97/137	6	6	7	7	6	6	6.33
97/138	5	5	5	5	5	5	5.00
97/139	5	5	6	6	5	5	5.33
97/140	7	7	8	8	7	7	7.33
97/141	6	4	5	5	5	5	5.33
97/142	4	4	5	5	4	5	4.33
97/143	5	5	5	5	4	4	4.66
97/144	3	3	4	4	3	3	3.33
97/145	6	4	5	5	5	5	5.33
97/146	6	6	6	6	7	6	6.33
97/147	6	6	5	5	4	4	5.00
97/148	5	4	5	4	5	4	5.00
97/149	4	4	4	4	4	5	4.00
97/150	3	4	3	3	3	3	3.00
97/151	4	4	4	4	4	4	4.00
97/152		4	4	4	4	3	4.00
97/153	5	5	6	5	4	4	5.00
97/154	5	5	5	5	5	5	5.00

97/155	6	7	8	8	7	6	7.00
97/156	7	6	5	5	6	5	6.00
97/157	8	7	5	5	5	6	6.00
97/158	7	7	7	7	6	6	6.66
97/159	5	5	6	5	6	6	5.66
97/160	6	5	5	5	6	6	5.66
97/161		4	3	4	3	3	3.00
97/162	5	5	6	6	5	5	5.33
97/163		4	3	3	3	3	3.00
97/164	3	3	3	2	3	3	3.00
97/165		5	3	3	3	3	3.00
97/166	7	6	7	6	6	6	6.66

APPENDIX 21

1997 INTAKE

KEY STAGE 3 RESULTS: NON CORE SUBJECTS (2000)

TEACHER ASSESSMENT LEVELS

DT = Design Technology Hi = History Geo = Geography Mu = Music

ICT = Information and Communications Technology

PE = Physical Education MFL = Modern Foreign Languages

	Hi	Geo	DT	ICT	MFL	Art	Mu	PE
97/1	5	3		4		A	A	A
97/2	3	3	5	5	3	C	A	A
97/3	7	7	6	6	6	C	C	C
97/4	6	7	8	4	6	B	C	B
97/5	5	6	5	4	5	C	B	B
97/6	7	7	7	5	6	B	C	B
97/7	5	4	5	5	5	C	C	B
97/8	6	5	6	5	5	B	C	C
97/9	6	6	7	6	5	C	B	B
97/10	6	7	7	5	6	C	C	B
97/11	3	5	5	4	3	C	C	B
97/12	6	6	7	6	5	C	C	B
97/13	6	6	5	5	4	C	C	B
97/14	4	5	5	5	4	C	C	B
97/15	4	4	5	4	4	B	A	B
97/16	7	7	7	6	5	C	C	B
97/17	6	7	7	6	5	C	C	B
97/18	5	5	5	5	4	B	A	B
97/19	5	6	5	5	5	B	B	B
97/20	6	6	6	6	6	C	C	B
97/21	5	4	5	5	4	C	A	B
97/22	6	5	5	5	4	C	C	B
97/23	5	6	6	4	5	C	C	B
97/24	6	6	7	4	4	B	C	B
97/25	4	4	4	5	4	B	A	B
97/26	5	5	5	4	5	C	C	A
97/27	7	7	7	6	6	C	D	C
97/28	6	6	4	5	4	C	B	B
97/29	4	4	4	5	4	B	A	A
97/30	4	3	5	6	3	B	A	A
97/31	4		5	5	4	B	A	B
97/32	4	4	5	5	4	B	A	B
97/33	5	5	4	4	4	B	A	A
97/34	5	5	6	5	5	B	C	B
97/35	5	6	5	5	4	C	D	B
97/36	4	3	6	4	4	B	B	A
97/37	5	5	6	5	3	B	B	B
97/38	4	3	5	6	3	C	B	A

97/39	5	4	5	4	4	C	B	B
97/40	6	7	7	4	6	C	C	B
97/41	6	6	7	5	5	C	C	C
97/42	4	4	5	6	4	C	B	B
97/43	4	4	5	4	4	B	B	B
97/44	5	5	5	5	4	C	B	B
97/45	6	4	5	6	4	C	B	B
97/46	5	6	5	4	4	C	B	B
97/47	8	7	7	6	6	C	B	B
97/48	6	7	7	6	6	C	D	B
97/49	4	4	5	4	4	B	B	B
97/50	5	5	6	5	4	C	B	B
97/51	6	6	7	7	6	C	C	B
97/52	4	3	5	5	4	C	A	B
97/53	4	5	5	5	4	C	B	B
97/54	6	5	6	5	4		B	B
97/55	5	4	5	5	4	C	B	B
97/56	3	5	4	4	3	C	A	B
97/57	4	4	4	5	3	B	A	B
97/58	7	7	7	6	6	C	D	C
97/59	5	5	5	5	6	C	A	B
97/60	4	4	5	6	4	C	B	B
97/61	7	6	6	6	6	C	B	A
97/62	6	6	6	5	6	C	A	B
97/63	6	4	5	5	4	C	A	B
97/64	5	5	4	5	4	C	B	A
97/65	6	5	5	5	4	C	C	B
97/66	6	6	6	6	4	C	B	B
97/67	4	5	5	5	4	B	A	B
97/68	5	3	6	5	4	C	B	A
97/69	5	5	7	5	5	B	B	B
97/70	5	6	5	3	4	B	B	A
97/71	7	7	7	5	6	C	B	B
97/72	4	3	5	5	3	B	A	B
97/73	4	5	4	4	4	B	A	B
97/74	4	4	5	5	4	B	A	A
97/75	5	6	6	6	5	C	C	B
97/76	4	3	5	5	4	B	A	B
97/77	6	5	6	5	5	C	D	B
97/78	6	7	6	4	5	C	D	B
97/79	5	4	6	4	4	C	B	B
97/80	5	5	6	4	4	B	B	B
97/81	6	7	7	4	5	B	C	B
97/82	6	7	6	6	6	C	C	B
97/83	5	5	5	5	4	C	B	B
97/84	5	4	5	4	4	B	C	B
97/85	4	5	6	5	4	B	A	B
97/86	7	7	7	6	6	C	B	B
97/87	6	7	6	6	5	C	C	B
97/88	6	6	6	5	5	C	A	B
97/89	5	6	7	5	5	C	C	A
97/90	5	6	5	5	4	C	C	B
97/91	6	7	7	4	6	C	C	B
97/92	6	5	7	5	6	C	C	B
97/93	6	7	7	5	6	C	D	C

97/94	5	5	5	5	5	C	C	A
97/95	6	7	6	5	5	C	B	B
97/96	6	5	6	4	5	C	D	C
97/97	3	3	5	5	3	A	A	A
97/98	5	4	6	5	3	C	B	A
97/99	3	4	4	4	3	C	C	B
97/100	6	7	6	6	6	C	D	B
97/101	5	3	5	5	4	C	A	B
97/102	5	4	4	5	3	C	A	B
97/103	7	7	7	5	6	C	C	B
97/104	5	6	6	5	4	C	B	B
97/105	5	5	6	5	4	B	C	B
97/106	5	4	5	4	4	B	C	A
97/107	4	3	5	5	4	B	A	B
97/108	5	6	7	6	4	C	A	B
97/109	4	4	5	5	3	A	A	B
97/110	4	3	5	4	2	B	A	A
97/111	6	6	7	5	5	C	D	B
97/112	5	4	5	5	4	B	B	B
97/113	6	6	7	4	4	B	A	B
97/114	5	4	5	4	4	C	A	A
97/115	5	4	6	5	4	B	A	B
97/116	7	7	7	5	6	C	C	B
97/117	7	7	6	6	7	C	C	B
97/118	4	3	5	4	4	A	A	B
97/119	5	3	4	5	5	B	B	B
97/120	4	5	4	4	3	B	A	B
97/121	4	4	4	4	3	C	A	A
97/122	6	7	4	5	6	C	D	B
97/123	4	3	4	5	4	C	A	B
97/124	6	7	7	5	5	C	D	C
97/125	6	6	7	5	4	B	A	B
97/126	4	4	5	5	4	C	B	B
97/127	6	6	7	6	5	C	C	B
97/128	6	7	7	6	5	B	C	B
97/129	3	3	5	4	3	A	A	A
97/130	5	6	6	5	5	C	C	B
97/131	7	7	6	6	6	C	B	B
97/132	6	6	6	5	5	C	C	B
97/133	5	6	6	5	5	C	B	B
97/134	4	4	3	5	4	A	A	B
97/135	5	5	5	4	4	B	A	B
97/136	4	4	4	4	3	B	A	A
97/137	6	7	6	6	5	B	B	B
97/138	5	4	5	5	4	C	B	B
97/139	5	5	5	5	5	B	B	C
97/140	7	7	7	6	6	C	C	B
97/141	5	6	6	5	5	B	B	B
97/142	4	6	6	4	4	C	A	A
97/143	5	5	5	5	4	B	B	B
97/144	3	4	5	5	3	C	A	B
97/145	5	5	5	5	4	B	B	B
97/146	6	6	6	6	5	C	B	B
97/147	5	5	5	6	4	B	B	B
97/148	5	6	4	4	4	C	A	A

97/149	5	4	5	5	4	B	A	B
97/150	4	3	4	3	2	B	A	B
97/151	4	3	4	4	3	C	A	A
97/152	4	4	4	5	2	C	A	B
97/153	4	4	5	5	4	C	B	B
97/154	5	4	5	5	4	B	A	B
97/155	6	7	7	5	6	B	A	B
97/156	5	6	7	4	5	C	C	B
97/157	7	7	6	5	6	C	D	B
97/158	7	7	7	5	6	C	D	B
97/159	6	6	5	6	4	C	B	B
97/160	5	6	5	4	4	C	B	B
97/161	3	3	5	4	3	A	A	C
97/162	4	6	6	5	5	C	B	B
97/163	4	3	4	4	3	A	B	B
97/164	3	3	4	4	2	A	A	A
97/165	3	4	4	5	4	B	A	B
97/166	6	5	6	6	5	C	C	B

APPENDIX 22

1998 INTAKE

RESULTS OF CAT

(LEVEL D, TAKEN IN 1998)

SS = Standard Age Score (Mean = 100)

	Verbal SS	Quantitative SS	Non-verbal SS	Average SS
98/1	104	110	110	108
98/2	92	106	122	107
98/3	77	71	76	75
98/4	94	104	98	99
98/5	110	114	109	111
98/6	74	76	89	80
98/7	78	78	77	78
98/8	88	85	97	90
98/9	85	90	89	88
98/10	89	93	110	97
98/11	96	89	94	93
98/12	110	105	125	113
98/13	88	100	93	94
98/14				
98/15	101	105	95	100
98/16	104	99	108	104
98/17	118	113	113	115
98/18	110	105	88	101
98/19	83	85	74	81
98/20	95	94	96	95
98/21	108	116	114	113
98/22	101	94	108	101
98/23	100	103	116	106
98/24	77	71	70	73
98/25	90	103	111	101
98/26	102	115	108	108
98/27	110	105	115	110
98/28	79	83	77	80
98/29	101	108	99	103
98/30				
98/31	80	80	71	77
98/32	120	116	108	115
98/33				
98/34	109	106	119	111
98/35	87	86	84	86
98/36	118	94	113	108
98/37	104	100	106	103
98/38	95	104	91	97
98/39	126	110	112	116
98/40	113	108	105	109
98/41	103	114	107	108
98/42	88	98	92	93

98/43	109	93	96	99
98/44	114	113	110	112
98/45	87	95	84	89
98/46	73	78	88	80
98/47	104	85	96	95
98/48	110	105	110	108
98/49	100	100	117	106
98/50	106	113	100	106
98/51	104	97	105	102
98/52	95	107	113	105
98/53	99	114	109	107
98/54	111	96	117	108
98/55	113	119	93	108
98/56	102	112	106	107
98/57				
98/58	114	96	115	108
98/59	87	83	74	81
98/60	103	105	108	105
98/61	112	110	106	109
98/62	123	113	101	112
98/63	117	108	110	112
98/64	117	98	92	102
98/65	74	78	92	81
98/66	109	106	117	111
98/67	89	88	92	90
98/68	93	99	99	97
98/69	92	100	88	93
98/70	115	101	110	109
98/71	97	96	98	97
98/72	72	84	87	81
98/73	98	81	84	88
98/74	90	83	96	90
98/75	105	115	110	110
98/76	98	88	80	89
98/77	120	110	114	115
98/78	107	100	91	99
98/79	85	87	87	86
98/80	110	92	97	100
98/81	116	104	110	110
98/82	95	99	92	95
98/83	116	103	98	106
98/84	105	106	107	106
98/85	88	92	99	93
98/86	99	87	83	90
98/87	91	93	79	88
98/88	98	81	87	89
98/89	118	98	115	110
98/90	90	107	99	99
98/91	120	108	112	113
98/92	78	81	95	85
98/93	108	93	110	104
98/94	109	89	100	99
98/95	130	119	108	119
98/96	108	126	114	116
98/97	89	92	98	93

98/98	90	115	97	101
98/99	101	95	109	102
98/100	90	85	83	86
98/101	98	108	89	98
98/102	104	95	97	99
98/103	87	93	94	91
98/104	86	89	94	90
98/105	70	71	80	74
98/106	92	84	93	90
98/107	109	113	106	109
98/108	104	107	113	108
98/109	96	109	113	106
98/110	108	96	103	102
98/111	97	96	86	93
98/112	94	102	109	102
98/113	101	91	109	100
98/114	119	106	109	111
98/115	112	106	103	107
98/116				
98/117	94	104	108	102
98/118	115	115	119	116
98/119	87	99	102	96
98/120	118	117	115	117
98/121	108	108	112	109
98/122	89	104	106	100
98/123	97	98	104	100
98/124	84	100	97	94
98/125	118	100	105	108
98/126	96	98	90	95
98/127	130	119	119	123
98/128	109	110	103	107
98/129	117	103	101	107
98/130	124	109	104	112
98/131	84	86	96	89
98/132	109	120	103	111
98/133	113	90	103	102
98/134	130	117	119	122
98/135	120	108	110	113
98/136	108	106	115	110
98/137	106	98	108	104
98/138	81	88	101	90
98/139				
98/140	83	77	76	79
98/141	104	102	97	101
98/142	79	86	86	84
98/143	103	114	130	116

APPENDIX 23

1998 INTAKE

KEY STAGE 2 RESULTS: CORE SUBJECTS (1998)

No teacher assessments were available for this intake year

	English Test	Maths Test	Science Test	Test Average
98/1	4	4	4	4.00
98/2		4	4	4.00
98/3	3	3	3	3.00
98/4	4	4	4	4.00
98/5	5	4	5	4.66
98/6	3	3	3	3.00
98/7	3	3	3	3.00
98/8	3	3	3	3.00
98/9	3	3	3	3.00
98/10	4	4	4	4.00
98/11	4	3	4	3.66
98/12	5	4	4	4.33
98/13	4	3	3	3.33
98/14				
98/15	4	4	5	4.33
98/16	5	4	5	4.66
98/17	5	5	4	4.66
98/18	4	4	3	3.66
98/19	4	3	3	3.33
98/20	4	4	4	4.00
98/21	4	4	5	4.33
98/22	4	3	4	3.66
98/23	4	5	5	4.66
98/24	3		3	3.00
98/25	4	4	5	4.33
98/26		4	4	4.00
98/27	4	4	4	4.00
98/28	2	3	3	2.66
98/29	4	4	4	4.00
98/30				
98/31	4	3	4	3.66
98/32	5	5	5	5.00
98/33	4	5	5	4.66
98/34	4	4	4	4.00
98/35				
98/36	5	3	5	4.33
98/37	4	4	4	4.00
98/38	4	3	4	3.66
98/39	5	5	5	5.00
98/40	5	4	4	4.33
98/41	4	5	4	4.33
98/42	4	4	3	3.66
98/43	4	4	5	4.33
98/44	4	4	5	4.33

98/45	4	4	4	4.00
98/46	3	3	3	3.00
98/47	3	3	4	3.33
98/48	5	4	5	4.66
98/49	4	3	3	3.33
98/50	4	4	4	4.00
98/51	4	4	4	4.00
98/52	5	4	5	4.66
98/53	4	4	4	4.00
98/54	4	4	4	4.00
98/55	5	5	4	4.66
98/56	5	4	4	4.33
98/57	4	3	3	3.33
98/58	5	4	5	4.66
98/59	4	3	3	3.33
98/60	4	4	4	4.00
98/61	4	4	4	4.00
98/62	5	4	5	4.66
98/63	5	4	4	4.33
98/64	5	4	4	4.33
98/65	3	4	4	3.66
98/66	5	4	5	4.66
98/67	4	4	4	4.00
98/68	5	4	4	4.33
98/69	4	3	3	3.33
98/70	5	5	5	5.00
98/71	3	3	4	3.33
98/72	4	4	4	4.00
98/73	3	3	4	3.33
98/74	4	3		3.50
98/75	4	5	5	4.66
98/76	4	3	3	3.33
98/77	5	5	5	5.00
98/78	5	4	4	4.33
98/79	4	3	3	3.33
98/80	4	4	4	4.00
98/81	5	4	4	4.33
98/82	4	3	3	3.33
98/83	4	3	4	3.66
98/84	4	4	4	4.00
98/85	4	4	4	4.00
98/86	4	3	3	3.33
98/87		4	4	4.00
98/88	4	3	4	3.66
98/89	4	4	4	4.00
98/90	4	4	4	4.00
98/91	4	5	5	4.66
98/92	3	3	3	3.00
98/93	5	4	5	4.66
98/94	4	4	4	4.00
98/95	5	5	5	5.00
98/96	5	4	4	4.33
98/97	4	3	4	3.66
98/98		3	4	3.50
98/99	4	3	4	3.66

98/100	4	4	4	4.00
98/101	4	4	4	4.00
98/102	4	4	4	4.00
98/103	3	3		3.00
98/104	4	3	4	3.66
98/105	3	3		3.00
98/106	4	3	4	3.66
98/107	5	4	5	4.66
98/108	4	4	4	4.00
98/109	4	4	4	4.00
98/110	5	4	4	4.33
98/111	4	3	4	3.66
98/112	4	3	4	3.66
98/113	4	3	4	3.66
98/114	5	4	4	4.33
98/115	5	4	5	4.66
98/116	4	3	3	3.33
98/117	4	4	4	4.00
98/118	4	5	5	4.66
98/119	4	3	4	3.66
98/120	4	4	4	4.00
98/121	5	4	5	4.66
98/122	4	4	5	4.33
98/123	4	4	3	3.66
98/124	3	3	3	3.00
98/125	4	4	4	4.00
98/126	4	4	4	4.00
98/127	5	5	5	5.00
98/128	5	4	4	4.33
98/129	5	4	4	4.33
98/130	5	5	5	5.00
98/131	4	3	3	3.33
98/132	5	5	4	4.66
98/133	4	4	4	4.00
98/134	5	5	4	4.66
98/135	5	5	5	5.00
98/136	5	4	4	4.33
98/137	5	4	4	4.33
98/138	3	3	4	3.33
98/139		5	4	4.50
98/140	3		4	3.50
98/141	4	4	4	4.00
98/142	3	3	3	3.00
98/143	4	5	4	4.33

APPENDIX 24

1998 INTAKE

KEY STAGE 3 RESULTS: CORE SUBJECTS (2001)

TEST AND TEACHER ASSESSMENT LEVELS

TA = Teacher assessment Eng = English Ma = Mathematics Sci = Science

	Eng Test	Eng TA	Ma Test	Ma TA	Sci Test	Sci TA	Test Average
98/1	6	5	6	6	6	5	6.00
98/2	6	5	6	6	5	5	5.66
98/3	5	5	3	3	3	3	3.66
98/4	6	5	6	6	6	6	6.00
98/5	6	6	7	6	7	7	6.66
98/6	5	5	4	4	4	4	4.33
98/7	5	5	5	5	5	4	5.00
98/8	4	4	4	4	4	4	4.00
98/9	5	5	4	4	4	4	4.33
98/10	5	5	5	5	5	6	5.00
98/11	5	5	4	4	4	4	4.33
98/12	6	7	7	7	6	6	6.33
98/13	6	5	6	6	5	4	5.66
98/14	8	8	7	7	7	3	7.33
98/15	6	5	5	5	5	5	5.33
98/16	5	5	6	6	5	6	5.33
98/17	6	6	7	7	6	6	6.33
98/18	5	5	5	5	5	5	5.00
98/19		4	4	4	4	4	4.00
98/20	7	6	5	5	5	5	5.66
98/21	5	6	7	7	6	6	6.00
98/22	5	5	6	6	5	5	5.33
98/23	6	6	7	7	6	6	6.33
98/24	4	4	3	3	4	3	3.66
98/25	5	5	6	6	6	6	5.66
98/26	6	5		6	5	6	5.50
98/27	6	6	6	6	6	5	6.00
98/28		3	3	3	2	4	2.50
98/29	6	5	4	4	4	4	4.66
98/30	6	5	3	3	4	4	4.33
98/31	3	4	3	3	3	4	3.00
98/32	8	8	7	7	6	6	7.00
98/33	7	7	7	7	6	6	6.66
98/34	8	7	7	6	6	6	7.00
98/35	5	5	4	4	5	4	4.66
98/36	7	7	6	6	7	6	6.66
98/37	6	5	6	6	6	6	6.00
98/38	5	4	5	5	4	5	4.66
98/39	7	6	7	7	6	6	6.66
98/40	7	6	7	7	6	6	6.66
98/41	6	6	7	7	6	6	6.33
98/42	5	5	6	6	5	4	5.33

98/43	6	5	6	6	6	6	6.00
98/44	6	7	7	7	7	6	6.66
98/45	5	5	4	4	4	3	4.33
98/46	4	4	3	3	4	3	3.66
98/47	6	5	5	4	5	5	5.33
98/48	7	6	7	6	6	6	6.66
98/49	6	5	6	6	5	5	5.66
98/50	5	4	5	5	5	4	5.00
98/51	6	5	6	6	6	6	6.00
98/52	5	5	6	6		5	5.50
98/53	7	6	7	7	5	6	6.33
98/54	5	6	6	6	6	5	5.66
98/55	6	6	6	6	5	6	5.66
98/56	6	6	6	6	5	6	5.66
98/57	6	5	4	4	4	4	4.66
98/58	6	5	6	6	6	6	6.00
98/59	5	4	4	4	4	4	4.33
98/60	5	6	6	6		6	5.50
98/61	7	6	6	6	6	4	6.33
98/62	6	6	7	7	6	6	6.33
98/63	7	7	7	7	6	5	6.66
98/64	5	6	6	6	6	6	5.66
98/65		3	5	5	3	3	4.00
98/66	6	7	6	6	6	6	6.00
98/67	5	5	4	5		4	4.50
98/68	6	5	6	6	6	5	6.00
98/69	5	4	5	5	5	5	5.00
98/70	8	7	7	7	7	7	7.33
98/71	5	5	5	5	5	6	5.00
98/72	4	4	3	3	3	3	3.33
98/73	6	5	5	5	5	5	5.33
98/74	3	5	4	4	4	4	3.66
98/75	6	6	7	7	6	6	6.33
98/76	5	4	4	4	4	4	4.33
98/77	7	7	7	7	7	6	7.00
98/78	5	5	5	5	5	5	5.00
98/79	4	5	4	4	4	4	4.00
98/80	5	5	6	6	5	5	5.33
98/81	8	7	7	7	7	7	7.33
98/82	5	4	5	5	4	4	4.66
98/83	5	6	5	5	6	6	5.33
98/84	6	6	6	6	6	5	6.00
98/85	5	5	6	6	5	6	5.33
98/86	6	6	5	5	5	4	5.33
98/87	5	5	5	5	5	5	5.00
98/88	6	6	6	6	6	5	6.00
98/89	5	6	6	6	6	6	5.66
98/90	5	5	4	4	4	5	4.33
98/91	7	7	7	7	6	6	6.66
98/92	4	4	4	4	4	5	4.00
98/93	6	6	6	6	6	5	6.00
98/94	6	5	5	5	6	5	5.66
98/95	6	7	6	7	5	6	5.66
98/96	6	6	7	7	6	6	6.33
98/97	5	5	6	6	5	5	5.33

98/98		4	5	5	4	4	4.50
98/99	5	6	6	6	5	5	5.33
98/100	5	5	4	4	5	4	4.66
98/101	5	4	5	5	5	5	5.00
98/102	7	6	6	6	6	6	6.33
98/103	3	4	4	4	3	4	3.33
98/104	5	5	5	5	5	5	5.00
98/105	4	5	3	4	3	3	3.33
98/106	4	4	3	4	4	5	3.66
98/107	6	6	6	6	6	6	6.00
98/108	4	5	5	5	5	5	4.66
98/109	5	4	6	6	4	4	5.00
98/110	6	6	6	6	6	6	6.00
98/111	4	5	5	5	5	4	4.66
98/112	5	6	6	6	5	6	5.33
98/113	4	5	5	5	5	5	4.66
98/114	7	7	7	7	7	6	7.00
98/115	6	6	7	7	6	6	6.33
98/116	5	4	5	5	4	5	4.66
98/117	7	7	7	7	7	6	7.00
98/118	6	7	8	8	6	6	6.66
98/119	4	4	4	4	4	4	4.00
98/120	7	7	7	7	6	7	6.66
98/121	5	5	6	6	5	5	5.33
98/122	4	5	6	6	5	4	5.00
98/123	4	5	6	6	4	6	4.66
98/124	4	5		4	4	4	4.00
98/125	6	7	6	6	6	7	6.00
98/126	5	5	5	5	5	5	5.00
98/127	8	7	8	8	7	7	7.66
98/128	8	7	6	6	6	5	6.66
98/129	7	7	5	6	6	6	6.00
98/130	8	7	7	7	7	7	7.33
98/131	5	5	5	5	3	4	4.33
98/132	7	6	7	7	6	6	6.66
98/133	5	5	5	5	4	4	4.66
98/134	7	7	7	7	6	6	6.66
98/135	6	6	8	8	7	7	7.00
98/136	7	6	6	6	6	6	6.33
98/137	6	6	6	6	6	5	6.00
98/138	5	5	5	5	5	6	5.00
98/139	6	7	6	7	6	6	6.00
98/140		3	3	3	3	4	3.00
98/141	5	5	5	5	6	5	5.33
98/142	4	4	4	4	3	4	3.66
98/143	5	6	6	7	6	6	5.66

APPENDIX 25

1998 INTAKE

TOTAL MARKS AND ENTRY LEVELS IN KEY STAGE 3 TESTS (2001)

There is only one level of entry in English

	English Mark	Maths Mark	Maths Level	Science Mark	Science Level
98/1	55	99	46	140	36
98/2	53	78	57	121	36
98/3	39	51	35	65	36
98/4	53	91	57	134	36
98/5	53	86	68	101	57
98/6	46	81	35	97	36
98/7	40	62	46	108	36
98/8	28	57	46	79	36
98/9	41	101	35	84	36
98/10	38	81	46	111	36
98/11	41	78	35	89	36
98/12	54	111	57	84	57
98/13	59	96	46	112	36
98/14	71	71	68	102	57
98/15	53	60	57	110	36
98/16	47	86	57	122	36
98/17	58	68	68	79	57
98/18	38	62	46	114	36
98/19		73	35	86	36
98/20	76	72	46	52	57
98/21	37	101	68	85	57
98/22	43	89	46	119	36
98/23	55	74	68	80	57
98/24	25	44	35	70	36
98/25	39	79	57	15	36
98/26					
98/27	53	90	57	145	36
98/28		35	35	38	36
98/29	55	55	46	88	36
98/30	53	40	35	73	36
98/31	18	44	35	66	36
98/32	81	99	68	93	57
98/33	70	124	57	85	57
98/34	86	81	68	94	57
98/35	38	99	35	110	36
98/36	82	66	57	102	57
98/37	54	95	46	132	36
98/38	40	78	46	95	36
98/39	70	69	68	85	57
98/40	71	98	57	76	57
98/41	53	75	68	93	57
98/42	42	89	57	102	36
98/43	64	99	46	137	36

98/44	62	74	100	101	57
98/45	37	56	46	80	36
98/46	21	57	35	75	36
98/47	60	62	46	113	36
98/48	81	67	68	86	57
98/49	55	101	46	119	36
98/50	37	58	57	103	36
98/51	60	102	46	133	36
98/52	44	94	46		
98/53	77	78	100	130	36
98/54	49	108	46	85	57
98/55	63	57	68	123	36
98/56	53	65	57	114	36
98/57	53	102	35	89	36
98/58	53	70	57	136	36
98/59	36	73	35	70	36
98/60	49	103	46		
98/61	80	84	57	84	57
98/62	63	66	68	87	57
98/63	75	87	68	15	36
98/64	40	102	46	131	36
98/65		59	46	54	36
98/66	60	88	57	136	36
98/67	44	83	35	124	
98/68	54	106	46	135	36
98/69	38	82	46	102	36
98/70	81	86	68	109	57
98/71	41		46	128	36
98/72	27	41	35	59	36
98/73	67	87	46	126	36
98/74	16		35	92	36
98/75	59	88	68	89	57
98/76	36	90	35	95	36
98/77	70	91	68	103	57
98/78	40	69	46	102	36
98/79	21	80	35	92	36
98/80	43	95	46	127	36
98/81	82	98	57	112	57
98/82					
98/83	39	88	46	81	57
98/84	54	121	46	82	57
98/85	46	104	46	115	36
98/86	58	80	46	112	36
98/87	42	77	46	130	36
98/88	59	90	46	143	36
98/89	41	124	46	147	36
98/90	36		46	98	36
98/91	78	93	68	93	57
98/92	23	72	35	96	36
98/93	60	77	57	134	36
98/94	65	86	46	138	36
98/95	62	65	100	69	57
98/96	54		68	96	57
98/97	39	89	46	122	36
98/98	11	60	46	77	36

98/99	41	115	46	130	36
98/100	40	88	35	109	36
98/101	36	71	46	102	36
98/102	72	85	57	135	36
98/103	16	81	35	63	36
98/104	43	67	46	103	36
98/105	28	63	35	58	36
98/106	22	63	35	93	36
98/107	56	53	68	133	36
98/108	30	86	46	110	36
98/109	38	93	46	100	36
98/110	58	87	57	76	57
98/111	23	69	46	107	36
98/112	36	95	46	67	57
98/113	27	58	46	104	36
98/114	70	108	57	101	57
98/115	61	107	57	98	57
98/116	36	60	46	94	36
98/117	80	116	57	110	57
98/118	53	107	100	93	57
98/119	21	98	35	81	36
98/120	76	81	100	93	57
98/121	37	89	57	127	36
98/122	24	91	46	110	36
98/123	25	96	46	100	36
98/124	22			95	36
98/125	62	91	57	96	57
98/126	40	65	46	111	36
98/127	77	111	68	122	57
98/128	72	76	57	135	36
98/129	76	87	46	84	57
98/130	88	106	100	113	57
98/131	37	67	46	68	36
98/132	70	88	100	80	57
98/133	36	63	46	97	36
98/134	70	75	68	92	57
98/135		121	100	107	57
98/136	81	79	57	75	57
98/137	61	69	57	144	36
98/138	38	71	46	118	36
98/139	59	62	68	77	57
98/140		34	35	50	36
98/141	37	87	46	138	36
98/142	26	87	35	62	36
98/143	39	64	68	90	57

APPENDIX 26

1998 INTAKE

KEY STAGE 3 RESULTS: NON CORE SUBJECTS (2001)

TEACHER ASSESSMENT LEVELS

DT = Design Technology Hi = History Geo = Geography Mu = Music

ICT = Information and Communications Technology

PE = Physical Education MFL = Modern Foreign Languages

	Hi	Geo	DT	ICT	MFL	Art	Mu	PE
98/1	5	5	5	5	5	6	5	5
98/2	4	5	5	5	5	6	5	4
98/3	4	3	4	4	4	4	4	4
98/4	6	5	7	5	5	5	5	5
98/5	5	5	7	4	6	7	7	6
98/6	6	5	5	4	4	7	6	5
98/7	5	5	5	5	5	5	5	5
98/8	4	3	5	4	4	6	5	4
98/9	4	4	4	5	5	5	5	5
98/10	5	5	7	5	6	6	6	5
98/11	4	4	5	5	5	6	5	5
98/12	6	6	7	6	7	7	7	5
98/13	5	5	5	5	6	5	5	6
98/14	7	6	7	6	7	7	6	5
98/15	5	5	5	5	5	6	4	5
98/16	5	4	6		6	6	5	5
98/17	6	6	6	5	6	6	8	6
98/18	5	4	6	5	4	6	6	5
98/19	3	4	4	4	4	4	4	4
98/20	7	6	7	5	6	7	5	
98/21	6	6	6	6	6	6	6	5
98/22	6	6	6	5	6	6	6	6
98/23	5	6	7	6	7	6	8	5
98/24	3	4	4	4	3	4	5	5
98/25	5	6	6	6	6	6	6	5
98/26	5	5	6	5	5	6	6	4
98/27	6	6	6	6	7	7	6	6
98/28	4	3	4	4	4	4	4	4
98/29	5	4	4	5	5	6	6	5
98/30	4	3	6	5	5		5	4
98/31	3	4	4	4	3	4	4	4
98/32	7	7	7	6	7	7	6	5
98/33	7	6	7	7	7	7	6	5
98/34	7	7	8	5	7	7	7	5
98/35	4	4	5	5	5	5	5	4
98/36	7	6	7	6	6	6	6	5
98/37	6	5	6		5	6	5	4
98/38	5	5	5	5	5	5	5	5

98/39	6	6	7	6	6	7	7	5
98/40	6	7	7	6	7	6	7	5
98/41	6	6	7	6	6	6	7	6
98/42	5	5	7	5	6	6	6	6
98/43	6	5	6	6	6	6	7	5
98/44	7	7	7	5	6	7	7	7
98/45	4	3	5	5	4	5	5	4
98/46	3	6	3	4	3	6	5	4
98/47	5	5	5	5	4	5	6	5
98/48	7	6	6	6	7	6	7	7
98/49	5	5	6	4	6	7	6	6
98/50	5	5	5	5	5	6	5	5
98/51	6	6	6	6	6	6	8	5
98/52	4	3	6		5	6	5	5
98/53	6	6	7	6	6	6	6	6
98/54	5	5	6	6	6	6	6	5
98/55	6	5	5	6	5	6	5	5
98/56	5	5		5	5	6	5	5
98/57	5	4	5	5	6	7	4	4
98/58	5	5	4	6	6	7	5	5
98/59	4	4	5	5	3	5	4	5
98/60	5	5	6	5	5	6	5	5
98/61	6	5	7	6	6	6	6	6
98/62	6	6	4	6	6	6	6	5
98/63	7	7	8	5	7	7	7	5
98/64	5	5	6	5	6	6	6	6
98/65	4	3	6		4	5	5	5
98/66	7	6	7	6	7	7	7	5
98/67	5	4	5	5	4	7	5	5
98/68	5	5	6	6	6	5	5	5
98/69	5	6	5	5	5	5	6	5
98/70	8	7	8	6	7	7	7	6
98/71	5	4	5	5	5	5	5	5
98/72	4	5	5	4	4	4	5	4
98/73	5	5	6	5	5	5	6	4
98/74	4	5	4	4	5	6	5	4
98/75	7	7	7	6	6	6	6	5
98/76	4	5	4		3	6	5	5
98/77	7	7	7	5	7	7	6	5
98/78	5	5	4	5	4	5	5	5
98/79	4	4	5	5	4	5	5	5
98/80	5	6	6	4	6	6	6	5
98/81	7	7	8	5	7	6	6	6
98/82	4	4	3	5	3	5	4	4
98/83	6	5	8	6	6	6	6	5
98/84	6	6	6	5	6	6	6	6
98/85	5	5	6	5	5	5	5	5
98/86	5	5	6	5	6	6	5	4
98/87	5	6	6		5	6	6	6
98/88	6	5	6	5	7	5	6	7
98/89	6	5	5	6	6	7	5	5
98/90	4	4	5	5	3		5	5
98/91	7	6	7	5	6	7	7	6
98/92	4	4	4	5	5	6	5	4
98/93	6	6	7	5	6	6	5	5

98/94	4	5	4	5	5	6	5	4
98/95	6	6	6		7	6	7	5
98/96	6	6	7	6	6	7	6	6
98/97	6	5	6	5	5	5	5	5
98/98	4	5	5	5	4	6	6	5
98/99	5	5	6		6	6	7	5
98/100	5	5	5	5	5	5	5	5
98/101	5	5	5	4	4	5	5	4
98/102	6	5	6	6	6	7	6	5
98/103	4	4	4	5	3	5	5	5
98/104	5	5	7	5	5	5	5	5
98/105	4	4	5	4	3	5	5	5
98/106	4	4	4	5	4	7	5	5
98/107	6	6	7	6	6	6	6	6
98/108	4	4	6	5	5	6	5	5
98/109	5	5	5	6	6	7	5	5
98/110	7	6	7	5	7	6	8	5
98/111	4	4	5		4	6	6	5
98/112	6	5	6	5	6	6	6	5
98/113	5	4	5	5	5	6	5	5
98/114	7	7	7	6	7	7	7	5
98/115	7	7	7	6	7		6	5
98/116	5	4	6		5	6	5	5
98/117	7	7	7	5	7	7	7	5
98/118	6	6	7	6	7	7	7	6
98/119	5	4	4	5	5	5	5	5
98/120	7	6	7	6	7	7	8	6
98/121	6	5	6	6	6	6	5	5
98/122	5	5	5	5	5	6	5	5
98/123	5	5	6	5	5	6	6	5
98/124	4	3	5	5	4	6	5	5
98/125	6	6	7	6	7	6	7	5
98/126	5	4	5	5	5		5	5
98/127	7	7	7	6	7	7	6	5
98/128	6	6	6	5	7	6	6	5
98/129	6	5	6	5	5	6	6	5
98/130	7	7	7	6	7	7	6	6
98/131	5	5	5	5	4	6	6	4
98/132	7	5		6	6	7	8	6
98/133	4	5	6	5	4	7	6	6
98/134	5	5	7	6	6	7	6	5
98/135	7	6	7		6	7	7	5
98/136	6	7	7	5	6	6	6	5
98/137	6	6	6	6	6	6	6	5
98/138	4	5	6		5	6	5	6
98/139	6	6	7	5	7	7	6	6
98/140	4	3	4	4	4	5	4	4
98/141	5	5	5	5	5	6	5	5
98/142	4	4	4	4	4	5	5	5
98/143	6	6	7	5	6	7	6	6

REFERENCES

Aristotle (1999) "Posterior Analytics, Book 1, 1-4, 31, and Book 2, 19" in Cooper, D.E. (ed) *Epistemology: The Classic Readings*, Oxford, Blackwell.

ATL (2000) *Response to DFEE Consultation on Proposals for Target Setting at Key Stage 3*, London, Association of Teachers and Lecturers.

Ball, S.J. (1998) "Educational Studies, Policy Entrepreneurship and Social Theory" in Slee, R., Weiner, G., and Tomlinson, S., (eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*, London, Falmer Press.

Barthes, R. (1974) *S/Z*,(tr. R. Miller), Oxford, Blackwell.

Barthes, R., (1993) *Mythologies*, (tr. A. Lavers), London, Vintage.

Bartlett, F.C. (1932) *Remembering*, Cambridge, Cambridge University Press.

Baudrillard, J. (1983) *Simulations*, (tr. P. Foss, P. Patton, and P. Beitchman), New York, Semiotext.

Berlak, H. (2000) "Cultural Politics, the Science of Assessment and Democratic Renewal of Public Education" in Filer, A. (ed), *Assessment: Social Practice and Social Product*, London, RoutledgeFalmer.

Black Paper (1977), London, Temple-Smith.

Black, P. (1986) "Assessment for Learning" in Nuttall, D.L. (ed), *Assessing Educational Achievement*, Lewes, Falmer Press.

Black, P. (2001) "Dreams, Strategies and Systems: portraits of assessment, past, present and future", *Assessment in Education*, vol. 8, no. 1, pp. 63-85.

Black P., and Wiliam D., (1998) Assessment and Classroom Learning, *Assessment in Education*, vol. 5, no.1, pp. 7-73.

Blair, A. (1996) *The Twentieth Anniversary Lecture*, given at Ruskin College Oxford, 16th December, 1996 (online) Ruskin College, Oxford, http://www.ruskin.ac.uk/archives/blairsp.htm. (accessed 15th February, 2002).

Broadfoot, P. (1996) "Assessment and Learning: Power or Partnership?" in Goldstein, H. and Lewis, T. (eds), *Assessment: Problems, Developments and Statistical Issues*, Ontario, Wiley and Sons.

Broadfoot, P. (1998) Records of Achievement and the Learning Society, *Assessment in Education*, vol. 5, no. 3, pp. 447-477.

Broadfoot, P. (2000) "Preface" in Filer, A. (ed), *Assessment: Social Practice and Social Product*, London, Routledge Falmer.

Broadfoot, P. and Pollard, A. (2000) "The Changing Discourse of Assessment Policy: The Case of English Primary Education" in Filer, A. (ed), *Assessment: Social Practice and Social Product*, London, RoutledgeFalmer.

Broadfoot, P. (2001) "Empowerment or Performativity? Assessment Policy in the late twentieth century" in Phillips, R. and Furlong, J. (eds), *Education, Reform and the State: Twenty Five Years of Politics, Policy and Practice*, London, Routledge Falmer.

Brown, D.E. (1991) *Human Universals*, London, McGraw-Hill.

Brown M., (1998) "The Tyranny of the International Horse Race" in Slee, R., Weiner, G., and Tomlinson, S., (eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements,* London, Falmer Press.

Brown M., McCallum B., Taggart B., Gipps C., (1997) The Validity of National Testing at age 11: the teachers view, *Assessment in Education* vol.4, no. 2, pp. 271-293.

Callaghan, J. (1976) Speech at Ruskin College, Oxford, 18th October, 1976. Manuscript in the Ruskin College Library.

Carr, W. and Kemmis, S. (1986) *Becoming Critical,* London, Falmer Press.

CEM (2002) *Standards over time,* (online), www.cem.dur.ac.uk, accessed 10th May 2002.

Chitty, C. (1994) "Consensus to conflict: the structure of educational decision-making transformed" in Scott D. (ed) *Accountability and Control in Educational Settings,* London, Cassell.

Chuang Tzu (1996) *The Book of Chuang Tzu,* (tr. M. Palmer), Harmondsworth, Penguin.

Clegg, A. (1976) "It's the same old story", *Times Educational Supplement,* 22nd October, 1976.

Cohen, G. (1993) "Everyday memory" in Cohen, G., Kiss, G., LeVoi, M., *Memory: Current Issues,* second edition, Buckingham, Open University Press.

Cohen, J. and Stewart, I. (1994) *The Collapse of Chaos: Discovering Simplicity in a Complex World*, New York, Penguin.

Cooper, D.E. (1999) *Epistemology: The Classic Readings*, Oxford, Blackwell.

Cooper, D. and Dunne, M. (2000) "Constructing the 'Legitimate' Goal of a 'Realistic' Maths Item: A Comparison of 10-11 and 13-14 year olds" in Filer, A. (ed), *Assessment: Social Practice and Social Product*, London, RoutledgeFalmer.

Coulby, D. (1989) "From Educational Partnership to Central Control", in Bash, L. and Coulby D., *The Education Reform Act: Competition and Control, London*, Cassell Educational.

Crooks T.J., Kane M.T., Cohen A.S., (1996) Threats to the valid use of assessments, *Assessment in Education*, vol. 3, no.3, pp.265-275.

Cresswell, M.J. (1996) "Defining, Setting and Maintaining Standards in Curriculum-embedded Examinations: Judgemental and Statistical Approaches, in Goldstein, H. and Lewis, T. (eds) *Assessment: Problems, Developments and Statistical Issues*, Ontario, Wiley and Sons.

Cumming J.S., Maxwell G.S., (1999) Contextualising Authentic Assessment, *Assessment in Education*, vol. 6, no. 2, pp.177-194.

Derrida, J. (1978) *Writing and Difference*, Bass, A. (trans), Athlone Press.

DFEE (1996) *Setting Targets to Raise Standards: A Survey of Good Practice*, London, DFEE.

DFEE (1997a) *From Targets to Action*, London, DFEE

DFEE (1997b) *1996 Study on Value-Added for 16-18 year olds in England*, Norwich, DFEE.

DFEE (1999) *National Learning Targets*, London, DFEE

DFEE (2000) *Consultation on proposals for target setting at Key Stage 3, London, DFEE.*

DFEE (2001) *Target setting at Key Stage 3 (KS3)*, London, DFEE.

DFES (2001) *Schools Achieving Success*, London, DFES.

Feldt, L.S. and Brennan, R.L. (1993) "Reliability" in Linn, R.L. (ed), *Educational Measurement*, 3rd edition, Phoenix AZ, Oryx Press.

Ferguson, G.A. and Takane, Y. (1989) *Statistical Analysis in Psychology and Education*, London, McGraw-Hill.

Fielding, T. (1998) "Target Setting: a case for caution", *Managing Schools Today*, vol. 8, no. 2, pp. 41-42.

Fielding, M. (1999) "Target Setting, Policy Pathology and Student Perspectives: Learning to labour in new times", *Cambridge Journal of Education*, vol. 29, no. 2, pp.277-287.

Filer, A. (2000a) "Socio-Historical and Cultural Contexts of Assessment Policy" " in Filer, A. (ed), Assessment: Social Practice and Social Product, London, Routledge Falmer.

Filer, A. (2000b) "Technologies of Testing" in Filer, A. (ed), *Assessment: Social Practice and Social Product*, London, Routledge Falmer.

Fitz-Gibbon C.T. (1996) *Monitoring Education: Indicators, Quality and Effectiveness*, London, Cassell.

Flecknoe, M. (2001) "Target Setting: Will it Help to Raise Achievement?", *Educational Management and Administration*, vol. 29, no. 2, pp. 217-228.

Foucault, M. (1979) *Discipline and Punish: The Birth of the Prison*, London, Peregrine Books.

Foucault, M., (1980) *Power/Knowledge: Selected Interviews and other writings, 1972-1977* (ed. C. Gordon), Harlow, Pearson Education Limited.

French, C.C. and Richards, A. (1993) "Clock this! An everyday example of a schema-driven error in memory", *British Journal of Psychology*, vol. 84, pp. 249-253.

Furlong, J. (2001) "Reforming teacher education, reforming teachers: Accountability, professionalism and competence" in Phillips, R. and Furlong, J. (eds), *Education, Reform and the State: Twenty Five Years of Politics, Policy and Practice*, London, Routledge Falmer.

Gardner, H. (1993, 2nd edn) *Frames of Mind: The Theory of Multiple Intelligences*, Fontana Press, London.

Gipps C. (1994) *Beyond Testing: Towards a Theory of educational assessment*, London, Falmer Press.

Goldstein, H. (2001) "Using Pupil Performance Data for Judging Schools and Teachers: scope and limitations", *British Educational Research Journal*, vol. 27, no.4, pp. 433-442.

Goldstein, H. and Lewis, T. (1996) "The Scope of Assessment" in Goldstein, H. and Lewis, T. (eds) *Assessment: Problems, Developments and Statistical Issues,* Ontario, Wiley and Sons.

Goldstein, H. and Sammons, P. (1997) "The Influence of Secondary and Junior Schools on Sixteen Year Examination Performance: A Cross-classified Multi-level Analysis", *School Effectiveness and School Improvement,* vol. 8, no. 2, pp.219-230.

Goleman, D., (1996) *Emotional Intelligence, Why it can matter more than IQ,* London, Bloomsbury.

Goleman, D. (1998) *Vital Lies, Simple Truths,* London, Bloomsbury

Gombrich, E.H. (1977) *Art and Illusion,* Fifth Edition, Oxford, Phaidon Press.

Gray, J. (1996) "The use of assessment to Compare Institutions", in Goldstein, H. and Lewis, T. (eds) *Assessment: Problems, Developments and Statistical Issues,* Ontario, Wiley and Sons.

Gribbin, J. (1995) *In Search of Schrodinger's Cat: Quantum Physics and Reality,* London, Transworld Publishers.

Guba, E.G. and Lincoln, Y.S. (1989) *Fourth Generation Evaluation,* London, Sage Publications.

Habermas, J. (1987) *Knowledge and Human Interests,* Cambridge, Polity Press.

Halstead M., (1994) "Accountability and Values" in Scott, D. (ed) *Accountability and Control in Educational Settings,* London, Cassell.

Hamilton, D. (1998) "The Idols of the Market Place" in Slee, R., Weiner, G., and Tomlinson, S., (eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements,* London, Falmer Press.

Hammersley, M (1999) "Taking Sides against Research: An Assessment of the Rationale for Partisanship", in Scott, D. (ed), *Values and Education,* London, University of London.

Hanson, F.A. (2000) "How Tests Create what they are Intended to Measure" in Filer, A. (ed), *Assessment: Social Practice and Social Product,* London, Routledge Falmer.

Husserl, E. (1964) *The Idea of Phenomenology,* (tr. W.P. Alston), The Hague, Hojhoff.

Kant, I. (1991) *Critique of Pure Reason,* London, Dent.

Kline, P. (1998) *The New Psychometrics: Science, Psychology and Measurement,* London, Routledge.

Koeslter, A. (1989) *The Ghost in the Machine,* London, Arkana (Penguin).

Koretz D., (1998) Large-scale Portfolio Assessments in the US: evidence pertaining to the quality of measurement, *Assessment in Education,* vol. 5, no. 3, pp. 309-334.

Kuhn, T.S. (1996) *The Structure of Scientific Revolutions,* Third Edition, Chicago, University of Chicago Press.

Lingard,B., Ludwig, J. and Luke, A. (1998) "School Effects in Postmodern Conditions" in Slee, R., Weiner, G., and Tomlinson, S., (eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*, London, Falmer Press.

Linn, R.L. (1993) "Current Perspectives and Future Directions" in Linn, R.L. (ed), *Educational Measurement*,3rd edition, Phoenix AZ, Oryx Press.

Locke, J. (1961) *An Essay Concerning Human Understanding*, Vol 1. (ed. J. Yolton), London, Dent.

Loftus, E. (1993) "The reality of repressed memories", *American Psychologist*, vol. 48, pp. 518-537.

Lyotard, J-F, (1984) *The Postmodern Condition: A Report on Knowledge*, (tr. G. Bennington and B. Massumi), Manchester, Manchester University Press.

Madaus, G.F. and Horn, C. (2000) "Testing Technology: the Need for Oversight" in Filer, A. (ed), *Assessment: Social Practice and Social Product*, London, Routledge Falmer.

Marx, K (1971) *A Contribution to the Critique of Political Economy, (trans. Ryazanskaya, S.W., ed. Dobb, M.)*, London, Lawrence and Wishart.

McCulloch, G. (2001) "The reinvention of teacher professionalism" in Phillips, R. and Furlong, J. (eds), *Education, Reform and the State: Twenty Five Years of Politics, Policy and Practice*, London, RoutledgeFalmer.

McCarty, C., Horleston, S. and Muijs, D. (2001) *Summary of an investigation into the performance on common items in English, Mathematics, Sciences and ICT, taken by pupils obtaining the same National Curriculum levels at Key Stages 1, 2 and 3.* Unpublished manuscript sent to the author.

McDonald, A.S. (2001) "The Prevalence and Effects of Test Anxiety in School Children", *Educational Psychology,* vol. 21, no. 1, pp.89-101.

Messick, S. (1993) "Validity", in Linn, R.L. (ed), *Educational Measurement,* Phoenix, AZ, Oryx Press.

Moody, I. (2001) "A case-study of the predictive validity and reliability of Key Stage 2 test results, and teacher assessments, as baseline data for target-setting and value-added at Key Stage 3", *The Curriculum Journal*, vol. 12, no. 1, pp. 81-101.

Morrison, H.G., Busch, J.C. and D'Arcy, J. (1994) "Setting Reliable National Curriculum Standards: a guide to the Anghoff procedure", *Assessment in Education,* vol.1, no.2, pp.181-199.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988) *School Matters: the Junior Years,* Wells: Open Books.

Mortimore, P. (1998) The Road to Improvement: *Reflections on School Effectiveness*, Lisse, Swets and Zeitlinger.

NFER (1986) *Cognitive Abilties Test Manual,* Windsor: NFER.

NFER (1998) *CAT to Key Stage 3 Indicators, Autumn 1998,* Windsor, NFER.

NFER (1999) *CAT to Key Stage 3 Indicators, Autumn 1999,* Windsor, NFER.

NFER (2000) *CAT to Key Stage 3 Indicators, Autumn 2000,* Windsor, NFER.

NFER (2001) *CAT2E to Key Stage 3 Indicators, Autumn 2001,* Windsor, NFER.

Nietzsche, F. (1999) "On Truth and Lies in a Nonmoral Sense" in Cooper, D.E. (ed) *Epistemology: The Classic Readings*, Oxford, Blackwell.

Nyaya-Sutras (1999) "Book 1, Chapter 1 and Book 2, Chapter 1, with Vatsyayana's Commentary" in Cooper, D.E. (ed) *Epistemology: The Classic Readings*, Oxford, Blackwell.

Perrenoud, P. (1991) "Towards a Pragmatic Approach to Formative Evaluation" in Weston, P. (ed), *Assessment of Pupil Achievement: Motivation and School Success,* Amsterdam, Swets and Zeitlinger.

Phillips, R. (2001) "Education, the state and the politics of reform: the historical context, 1976-2001" in Phillips, R. and Furlong, J. (eds), *Education, Reform and the State: Twenty Five Years of Politics, Policy and Practice,* London, Routledge Falmer.

Plato (1993) *Republic,* (tr. R. Watefield), Oxford, OUP.

Popper, K.R. (1972) *Conjectures and Refutations,* Fourth edition, London, Routledge and Kegan Paul.

Pryor, J. and Torrance, H. (2000) "Questioning the Three Bears: The Social Construction of Classroom Assessment" in Filer, A. (ed) *Assessment: Social Practice and Social Product,* London, RoutledgeFalmer.

QCA (1998) *Monitoring Consistency in Teacher Assessment and the Impact of SCAA's guidance material at Key Stages 1, 2, and 3: Final Report,* London, QCA.

Rea, J. and Weiner, G. (1998) "Cultures of Blame and Redemption – When Empowerment Becomes Control: Practitioners' Views of the Effective Schools Movement" in Slee, R., Weiner, G., and Tomlinson, S., (eds) School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements, London, Falmer Press.

Reay, D. and Wiliam, D. (1999) "'I'll be nothing': structure, agency and the construction of identity through assessment", *British Education Research Journal,* vol. 25, no. 3, pp. 343-354.

Riddell, S., Brown, S. and Duffield, J. (1998) "The Utility of Qualitative Research for Influencing Policy and Practice in School Effectiveness" in Slee, R., Weiner, G., and Tomlinson, S., (eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements,* London, Falmer Press.

Russell, B. (1979) *History of Western Philosophy,* London, Unwin.

Rutter, M., Maughan, B., Mortimore, P., and Ouston, J. (1979) *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children,* London, Open Books.

Sadler D.R. (1998) Formative assessment: revisiting the territory, *Assessment in Education,* vol. 5, no. 1, pp. 77-84.

Sammons P., Thomas S., Mortimore P., (1997) *Forging Links: Effective Schools and Effective Departments,* London, Paul Chapman Publishing.

Saunders, L. (1999) "A Brief History of Educational 'Value-Added': How Did We Get To Where We Are?", *School Effectiveness and School Improvement,* vol. 10, no. 2, pp. 233-256.

Saunders, L. (2000) "Understanding school's use of 'value-added' data: the psychology and sociology of numbers", *Research Papers in Education,* vol. 15, no.3, pp.241-258

SCAA (1994) *Value-added Performance Indicators for Schools,* London, SCAA.

SCAA (1997a) *The Value-added National Project: Final Report,* Middlesex, SCAA.

SCAA (1997b) *Target Setting and Benchmarking in Schools,* Middlesex, SCAA.

SCAA (1997c) *Making Effective use of Key Stage 2 Assessments,* London, SCAA.

Schofield, J.W. (1993) "Increasing the generalisability of qualitative research", in Hammersley, M. (ed.), *Educational Research: current issues,* London, Paul Chapman Publishing.

Scott, D. (1994a) "Making Schools Accountable: assessment policy and the Education Reform Act", in Scott, D. (ed), *Accountability and Control in Educational Settings,* London, Cassell.

Scott, D. (1994b) "Afterword" in Scott, D. (ed), *Accountability and Control in Educational Settings*, London, Cassell.

Scott, D. (1999) "Endpiece: Reaching an Agreement about Rationality", in Scott, D. (ed), *Values and Education*, London, University of London.

Scott, D. (2000) *Realism and Educational Research: New Perspectives and Possibilities*, London, RoutledgeFalmer.

Sebatane E.M. (1998) Assessment and classroom learning: a response to Black and Wiliam, *Assessment in Education*, vol. 5, no. 1, pp. 123-130.

Segall, M.H., Campbell, D.T. and Herskovitz, M.J., (1963) "Cultural differences in the perception of geometric illusions", *Science*, vol. 139, pp. 769-771.

Sextus Empiricus (1999) "Outlines of Pyrrhonism" in Cooper, D.E. (ed) *Epistemology: The Classic Readings*, Oxford, Blackwell.

Skinner, B.F. (1938) *The Behaviour of Organisms*, New York, Appleton-Century-Crofts.

Skinner, B.F. (1968) *The Technology of Teaching*, New York, Appleton-Century-Crofts.

Slee, R. and Weiner, G. (1998) "Introduction: School Effectiveness for Whom?" in Slee, R., Weiner, G., and Tomlinson, S., (eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*, London, Falmer Press.

TES (1976) "Great debate opens with an anticlimax", *Times Educational Supplement*, 22nd October, 1976.

TES (2002) "Heads aiming at test targets", *Times Educational Supplement,* June 14, 2002.

TGAT (1988) *National Curriculum Task Group on Assessment and Testing,* London, DES.

Tikly C., Noss R., Goldstein H., (2000) "Pupils and Policy: what is the target?", *Forum (For Promoting 3-19 Comprehensive Education),* vol. 42, no. 1, pp. 12-14.

Torrance, H. (2000) "Postmodernism and Educational Assessment" in Filer, A. (ed), *Assessment: Social Practice and Social Product,* London, Routledge Falmer.

Tymms, P., FitzGibbon, C. (2001) "Standards, achievement and educational performance: a cause for celebration" in Phillips, R. and Furlong, J. (eds), *Education, Reform and the State: Twenty Five Years of Politics, Policy and Practice,* London, Routledge Falmer.

Turnbull, C.M. (1961) *The Forest People,* New York, Simon and Schuster.

Usher, R. (1999) "Overstepping the Limits: Postmodernism and the Research Process", in Scott, D. (ed), *Values and Education,* London, University of London.

Walsh, P. (1999) "Values and Detachment in Educational Research: The Need for Nuance", in Scott, D. (ed), *Values and Education,* London, University of London.

Watson J.B. (1930) *Behaviourism,* New York, Norton.

Wiggins, A. and Tymms, P. (2000) *Dysfunctional Effects of Public Performance Indicator Systems: A comparison between English and Scottish Primary Schools*, Paper presented to the European Conference in Educational Research, September 2000, (online), www.leeds.ac.uk/educol/documents/0000 1555.doc.

Wiliam, D. (1993) "Validity, dependability and reliability in National Curriculum assessment", *The Curriculum Journal*, vol. 4, no. 3, pp. 335-350.

Wiliam, D. (1995) "Combination, Aggregation and Reconciliation: evidential and consequential bases", *Assessment in Education*, vol. 2, no. 1, pp. 53-73.

Wiliam, D. (1996) "Meanings and Consequences in Standard Setting", *Assessment in Education*, vol.3, no. 3, pp.287-307.

Wiliam, D. (2001) "An overview of the Relationship between Assessment and the Curriculum" in Scott, D. (ed) *Curriculum and Assessment*, London, Ablex Publishing.

Wittgenstein, L. (1969) *On Certainty*, (tr. D. Paul and G.E.M. Anscombe), Oxford, Blackwell.

Wolf, A.(1998) Assessment as National Policy: The National Council for Vocational Qualifications and its quest for a pedagogical revolution, *Assessment in Education*, vol. 5, no. 3, pp. 413-445.

Wood, D. (1998) *How Children Think and Learn: The Social Contexts of Cognitive Development*, Oxford, Blackwell.

Woodhouse, G. and Goldstein, H. (1996) "The Statistical Analysis of Institution-based Data" in Goldstein, H. and Lewis, T. (eds), *Assessment: Problems, Developments and Statistical Issues*, Ontario, Wiley and Sons.

Young, M.F.D. (1971) *Knowledge and Control*, London, Collier-Macmillan.